ence Library

JUL 5 1945

Vol. 35, No. 3 June, 1945

Journal of Experimental Psychology

EDITED BY

SAMUEL W. FERNBERGER, ON LEAVE
FRANCIS W. IRWIN, ACTING EDITOR
UNIVERSITY OF PENNSYLVANIA

CONSULTING EDITORS

CHARLES W. BRAY, *Princeton University;* ELMER K. CULLER, *University of Rochester;* CLARENCE H. GRAHAM, *Brown University;* JOY P. GUILFORD, *University of Southern California;* FRANCIS W. IRWIN, *University of Pennsylvania;* DONALD G. MARQUIS, *Yale University;* ARTHUR W. MELTON, *University of Missouri;* CARROLL C. PRATT, *Rutgers University.*

CONTENTS

The Effect of Satiation on the Behavior Mediated by a Habit of Maximum Strength: S. KOCH AND W. J. DANIEL 167

Drive Interaction: II. Experimental Analysis of the Role of Drive in Learning Theory: H. H. KENDLER 188

The Role of Absolute Initial Response Strength in Simple Trial-and-Error Learning: P. S. SIEGEL 199

Affective Distance and Other Factors Determining Reaction Time in Judgments of Color Preference: W. C. SHIPLEY, J. I. COFFIN, AND K. C. HADSELL 206

An Experimental Analysis of Dynamic and Static Equilibrium: R. C. TRAVIS 216

Apparent Whiteness in Relation to Albedo and Illumination: R. E. TAUBMAN 235

Selective Forgetting when the Subject is not 'Ego-Involved': F. J. SHAW AND A. SPOONER 242

Autonomic Response in Posthypnotic Amnesia: M. E. BITTERMAN AND F. L. MARCUSE 248

PUBLISHED BI-MONTHLY,
BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
PRINCE AND LEMON STS., LANCASTER, PA.
AND NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879

PUBLICATIONS OF

THE AMERICAN PSYCHOLOGICAL ASSOCIATION

WILLARD L. VALENTINE, *Business Manager*

PSYCHOLOGICAL REVIEW

HERBERT S. LANGFELD, *Editor*
Princeton University

Contains original contributions only, appears bi-monthly, January, March, May, July, September, and November, the six numbers comprising a volume of about 540 pages.
Subscription: $5.50 (Foreign, $5.75). Single copies, $1.00.

PSYCHOLOGICAL BULLETIN

JOHN E. ANDERSON, *Editor*
University of Minnesota

Contains critical reviews of books and articles, psychological news and notes, university notices, and announcements. Appears monthly (10 issues), the annual volume comprising about 665 pages. Special issues of the BULLETIN consist of general reviews of recent work in some department of psychology.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

JOURNAL OF EXPERIMENTAL PSYCHOLOGY

S. W. FERNBERGER, *Editor, on Leave,* FRANCIS W. IRWIN, *Acting Editor*
University of Pennsylvania

Contains original contributions of an experimental character. Appears bi-monthly (beginning 1944), one volume per year, each volume of six numbers containing about 520 pages.
Subscription: $7.00 (Foreign, $7.25). Single copies, $1.25.

PSYCHOLOGICAL ABSTRACTS

WALTER S. HUNTER, *Editor*
Brown University

Appears monthly, the twelve numbers and an index supplement making a volume of about 700 pages. The journal is devoted to the publication of non-critical abstracts of the world's literature in psychology and closely related subjects.
Subscription: $7.00 (Foreign, $7.25). Single copies, 75c.

PSYCHOLOGICAL MONOGRAPHS

JOHN F. DASHIELL, *Editor*
University of North Carolina

Consist of longer researches or treatises or collections of laboratory studies which it is important to publish promptly and as units. The price of single numbers varies according to their size. The MONOGRAPHS appear at irregular intervals and are gathered into volumes of about 350 pages.
Subscription: $6.00 per volume (Foreign, $6.30).

JOURNAL OF ABNORMAL AND SOCIAL PSYCHOLOGY

GORDON W. ALLPORT, *Editor*
Harvard University

Appears quarterly, January, April, July, October, the four numbers comprising a volume of 400 pages. The journal contains original contributions in the field of abnormal and social psychology, reviews, notes and news.
Subscription: $5.00 (Foreign, $5.25). Single copies, $1.50.

JOURNAL OF APPLIED PSYCHOLOGY

DONALD G. PATERSON, *Editor*
University of Minnesota

Covers the applications of psychology in business, industry, education, etc. Appears bi-monthly, February, April, June, August, October, and December.
Subscription: $6.00. Single copies, $1.25.

Subscriptions, orders, and business communications should be sent to

THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

NORTHWESTERN UNIVERSITY, EVANSTON, ILLINOIS

Journal of
Experimental Psychology

VOL. 35, NO. 3 JUNE, 1945

THE EFFECT OF SATIATION ON THE BEHAVIOR MEDIATED BY A HABIT OF MAXIMUM STRENGTH

BY SIGMUND KOCH

Duke University

AND

WILLIAM J. DANIEL

University of North Carolina

INTRODUCTION

All theories of learning make a distinction, in some form, between habit structures and motivation. A major question facing any systematic theory of learning is, therefore: how do the processes represented by these variables combine in the determination of behavior? In other words, any behavior theorist must know the laws defining the joint influence of previously learned habit structures and present strength of motivation on the strength of the resultant behavior. An important aspect of this complex relationship is the problem of what happens to behavior when the strength of a previously learned habit is maximal, and the intensity of the ongoing motivation is zero. If a reaction is learned to the limit of practice in a given situation under strong and constant motivation, and if the motivation is now reduced to the neighborhood of zero, what variation will occur in the strength of the reaction? Although it should not be unduly difficult to approximate the conditions of this problem in experimental terms, speculation on this point has thus far outdistanced careful experimentation. The present experiment seeks to determine this relationship in as rigorous a way as extant techniques will permit.

In spite of the availability of a fair amount of relevant experimental evidence (**3**, **5**, **11**, **14**), psychological theorists have made widely divergent assumptions with regard to the question of how

strong habits are affected when the motivation of original learning disappears. At one extreme, we find the proponents of such principles as 'the transformation of mechanisms into drives' (Woodworth, **13**), 'functional autonomy of motives' (Allport, **1**), 'externalization of drive' (Anderson, **2**), etc. These theorists insist that strongly learned habits may acquire intrinsic motivational properties. At the other extreme, we find the emphasis of field theorists like Lewin (**7**) on the necessity of keeping structural concepts such as habit, and dynamic concepts such as motivation, clearly separated. On this view, behavior can never be derived from habit structures ('cognitive structure') alone; rather behavior can only be derived from 'psychological forces' which are jointly determined by habitual and motivational variables (i.e., cognitive structure and 'tension systems'). This would imply that the strength of behavior mediated by a habit of maximum strength [1] should fall to *zero*, after removal of the original learning motivation.

The first attempt to determine the joint influence of habit structures and motivation on behavior, in rigorous, quantitative fashion, may be found in Perin's experiment: 'Behavior Potentiality as a Joint Function of the Amount of Training and the Degree of Hunger at the Time of Extinction' (**8**). Since Perin's experiment seeks to define the relations among these variables in a perfectly general way, one aspect of his results provides a first, quantitative estimate of behavior strength, when strength of habit is maximal and motivation is zero. The present experiment thus takes Perin's investigation as its point of departure. For this reason, it will be necessary to describe Perin's experiment in some detail.

Working within the context of Hull's theory of behavior, Perin attempted to define the multiple functional relationship between 'habit strength' and 'drive intensity' on the one hand, and 'reaction potential' on the other.[2] In the terms of the experiment, habit strength corresponded to the number of previous reinforcements of a Skinner-box bar-depression reaction, drive intensity was varied by means of differing periods of food privation after satiation, and reaction potential was measured in three ways: (1) number of responses

[1] Field theorists would probably consider the notion of varying degrees of habit *strength* inadmissible. A similar purpose, however, seems to be served in Lewin's system by the concept of varying degrees of 'cognitive clarity' (**7**).

[2] Technical definitions of the concepts 'habit strength' and 'reaction potential' may be found in Hull's recent *Principles of behavior* (**6**). Perin's actual report is couched in the theoretically neutral terms of the laboratory operations which corresponded to 'habit strength' and 'drive intensity' in his experiment. These were 'amount of training' and 'degree of hunger,' respectively. The present summary of Perin's experiment is formulated in terms of its functional status within Hull's system as an exploration of the relation of habit strength and drive intensity to reaction potential. This departure from Perin's mode of presentation is made in order to bring out the theoretical significance of the experiment.

preceding a five min. extinction criterion, (2) time preceding the criterion, and (3) latency of the first three extinction responses. Three large groups of rats (*A*, *B*, and *C*) were given training in the Skinner-box bar-depression reaction. Each group consisted of four equivalent sub-groups. Habit strength varied for the different groups as follows: the sub-groups of *A* received 5, 8, 30, and 70 reinforcements, respectively; all four sub-groups of *B* received 16 reinforcements; and the *C* sub-groups were given 5, 10, 30, and 90 reinforcements, respectively.[3] After this training, the bar-depression reaction was extinguished in all groups to a five min. no response criterion. Drive intensity at the time of extinction varied for the different groups in the following way: the four sub-groups of *A* were all extinguished after three hours of hunger; the sub-groups of *B* were extinguished after 1, 3, 16, and 23 hours, respectively; and the *C* sub-groups were all extinguished after 22 hours of starvation.

From these data, Perin was able to determine the following relationships: (1) reaction potential as a function of habit strength when drive intensity is constant at three hours (group *A*) and at 22 hours (group *C*), and (2) reaction potential as a function of drive intensity when the number of reinforcements is held constant (group *C*). These relationships were first plotted empirically, and then rational equations describing the mathematical form of the functions were derived by appropriate curve fitting techniques. The final step was to combine the various functions into a single equation expressing reaction potential as a joint function of habit strength and drive intensity.

The result of this elaborate sequence of manipulations may be verbally paraphrased as follows: reaction potential increases as a positive growth function of habit strength, the asymptote of the growth function being in turn a slightly positively accelerated function of drive intensity. [In general, reaction potential turns out to be a *multiplicative* function of habit strength and drive intensity.] The aspect of Perin's results, however, of primary concern for the present problem is the following implication of his equations: even when drive intensity is zero, reaction potential remains a positive growth function of habit strength, approximating an asymptote of substantial magnitude. When drive intensity is zero and habit strength is *maximal*, reaction potential (as measured by resistance to extinction) attains a value as great as 28 percent of the reaction potential when *both* drive intensity and habit strength are maximal. When reaction potential is measured by latency, this figure jumps to 44 percent. A third measure of reaction potential—time preceding the attainment of the extinction criterion—yields a value of 50 percent.

[3] The data for group *C* were derived from an experiment performed by Williams (12). Since Williams' results constituted almost a third of Perin's data, it would be fair to regard the experiment as the 'Perin-Williams experiment.'

These figures seem surprisingly high. They seem quite out of line with expectations based on *any* theory which recognizes the full importance of motivation to the dynamics of behavior. An initial skepticism toward these figures is further justified by the fact that all of Perin's reaction potential values for zero drive are arrived at by extrapolation. Not a single group of animals was actually run under conditions of satiation. The closest empirically determined approximations to the zero drive condition were the four *C* sub-groups, extinguished under three hours of hunger and after 5, 8, 30 and 70 reinforcements respectively, and a single sub-group of *B*, extinguished under one hour of hunger and after 16 reinforcements. Thus, it appears desirable to check Perin's extrapolated reaction potential values for zero drive by direct experimental measurement.

The importance of testing this aspect of Perin's results is enhanced by the fact that Hull has based one of the most important postulates of his theory of behavior on Perin's work. Postulate 7 in Hull's *Principles of behavior* (**6**, 253)—which asserts that reaction potential is a multiplicative function of habit strength and drive intensity—is actually a restatement of Perin's results.[4] The quantitative formulation of the postulate is expressed by the equations derived in Perin's experiment. This postulate functions as the key assumption in Hull's present, highly ingenious analysis of 'primary motivation.' If Perin's extrapolated values for reaction potential under zero drive should turn out to be inaccurate, it becomes imperative that the quantitative formulation of the postulate be revised. It should be emphasized, of course, that a test of the zero drive points does not constitute a test of the *entire* set of relationships encompassed within the Perin-Hull reaction potential function. Both obtained and extrapolated reaction potential values at higher intensities of drive may still remain accurate, even if the zero drive values should turn out to be inaccurate. Nevertheless, any error in Perin's zero drive values would mean that postulate 7, in its present form, is capable of mediating a large number of false derivations within Hull's system.

In summary, then, the present experiment seeks to measure the strength of an instrumental behavior tendency brought to the limit of practice under strong motivation, when the original learning motivation is reduced to zero. A first, quantitative estimate of this value is

[4] Although Perin's results seem strongly to suggest that well-learned habits may acquire intrinsic motivational properties, it would be unfair to give the impression that Hull accepts this apparent implication. In **6**, p. 248 Hull indicates that any reaction potential in evidence after the drive of original learning has been eliminated may be due to 'residual' drives of an uncontrolled character present both during original learning and after elimination of the dominant need. Thus, Hull's position should not be confused with that of psychologists like Woodworth and Allport who impute intrinsic motivational properties to strong habits. Recent formulations of Hull's system emphasize the importance of motivation in all behavior dynamics just as strongly as do field theorists like Lewin.

provided by Perin, who measured the reaction potential of a Skinner-box bar-depression reaction under varying amounts of previous training and concurrent motivation. Since Perin's procedure seems admirably to realize the conditions of our problem, and since his zero drive results seem open to question, the design of our experiment consists in the application of Perin's procedure to the measurement of reaction potential after maximum reinforcement had been attained in the presence of strong hunger, but with concurrent hunger satiated as completely as possible. Such an experiment has a double significance: (1) it represents a test of the thesis that strong habits may acquire intrinsic motivational properties, and (2) it serves as a partial check on Hull's key postulate on primary motivation which embodies the results of the Perin experiment.

Subjects and Apparatus

The entire plan of the present experiment—including selection of *Ss*, apparatus and procedure—was designed to duplicate the conditions of Perin's experiment as precisely as possible.

The experimental group consisted of 32 rats selected from the colony of the University of North Carolina laboratory. These rats were approximately 90 days old at the beginning of the experiment, and they varied in weight from 133 to 255 gm. Four additional rats were used for preliminary standardization purposes, and 11 animals were thrown out during the course of the experiment for failure to satisfy the experimental conditions.[5]

The apparatus was a modification of the Skinner-box of the variety used by Ellson (4), Williams (12), and Perin (8, 9). This apparatus was obtained from the Yale laboratory, and so constitutes an exact replica of one of the four boxes used by Perin. The box consists of a 12 in. cubical experimental compartment enclosed in a large exterior shell which is packed with sound insulating material. The experimental compartment is ventilated by an air line blowing a gentle stream of compressed air into the apparatus, and it contains in one corner a bottle in which water is available throughout the experiment. On one wall of this compartment is a brass plate containing a horizontal slot through which a 55 mm. brass bar protrudes. Toward the bottom of the plate, and centered under the bar, is a small aperture through which pellets of food may fall into a food tray. During training, the apparatus was adjusted so that a pressure of 15 gm. or more exerted on the bar activated an electromagnetic food delivery device which released a single pellet of food into the food tray. At the same time, the bar depression activated: (1) a signal marker attached to a waxed paper polygraph, and (2) an impulse counter which was enclosed in a partially sound-proofed box. In this way, a permanent record was obtained of the occurrence and latency of each response, and *E* could conveniently read the ordinal number of each response from the impulse counter. A second signal marker on the polygraph was in circuit with a hand switch mounted on top of the box. This was used in order to record all time intervals during which the animals drank from the water bottle.

The pellets were made from a mixture of finely ground Purina checkers, whole wheat flour and water, according to a procedure suggested by Hull[6] and described by Siegal (10). Each pellet was a small, hard cylinder of this mixture, weighing approximately .075 gm.

During the extinction period, both the food delivery mechanism and the signal marker in circuit with the horizontal bar were disconnected. Thus, the only effect of bar depression was the activation of the impulse counter. The duration of the extinction period, until the attainment of the five min. no response criterion, was timed with an electric timer.

---

[5] The two main reasons for discarding the animals were: (1) failure of rats to learn the bar-depression reaction within a time comparable to that of Perin's groups, and (2) accidental malfunctioning of the apparatus during training, such as an occasional clogging of pellets in the food delivery tube.

[6] Personal communication from Clark L. Hull to William J. Daniel.

PROCEDURE

The two chief experimental conditions that our procedure was calculated to realize were: (1) to reinforce the bar-depression response to a point approaching the limit of habit strength, under conditions of strong hunger, and (2) to extinguish the response immediately after establishing complete hunger satiation. In achieving these conditions, an experimental routine similar to Perin's was followed. This routine may be divided into four phases: (1) establishing a feeding rhythm, (2) habituating the animals to the apparatus, (3) training at the bar-depression response, and (4) satiation and extinction.

(1) *Feeding rhythm.*—All animals were fed moist food in their home cages for one-half hour on each of seven successive days. Each rat was fed at the same time of day at which he was later experimented upon. The food consisted of a mixture of finely ground Purina checkers, whole wheat flour, and water. On days 8 and 9, the rats were fed for one-half hour, and at the customary time of day, in the apparatus. At these occasions, the food consisted of the standard pellets, a large quantity of which were placed in the food tray.

(2) *Habituation.*—The purpose of habituation was to establish a connection in the animals between the characteristic thump of the food delivery device, and the tendency to secure and eat the resulting pellet of food. Thus, on day 10, the rats were fed 40 pellets which were caused individually to appear in the food tray by *E* operating a hand switch mounted on top of the apparatus. When the animal was placed in the apparatus at the customary feeding time, three pellets were present in the food tray. Immediately after the rat consumed the third pellet, *E* released another pellet by closing the switch, and this was repeated until 20 pellets had been eaten. After the twentieth pellet had been received, the rat was returned to its living cage for 15 min., and was then reintroduced into the apparatus where the same procedure was followed for another 20 pellets.

(3) *Training.*—On day 11, the apparatus was adjusted so that it was necessary for the rat to depress the lever in order to secure a pellet of food, and the attempt was made to build up the bar-depression habit to a strength approaching the limit of learning. The maximum number of reinforcements given the animals in the Perin experiment was 70 in the case of group *A*, and 90 in the case of group *C*. According to Perin's curves (8, p. 101) for the 'number of reactions to produce extinction' as a function of 'number of reinforcement repetitions,' both 70 and 90 reinforcements eventuate in habit strengths close to the limit of learning. At 90 reinforcements, the curves are virtually flat, and the acceleration from 70 to 90 reinforcements is negligible.[7] Thus, it was decided to give the major part of the experimental group 70 reinforcements, while the possibility that this number did not sufficiently approximate the limit of habit strength was checked by giving the remainder of the animals 90 reinforcements. Accordingly, 26 of the animals were allowed 70 reinforced responses during the training period, and 6 rats were allowed 90 reinforcements.

The details of the training procedure were as follows. Before each rat was placed in the experimental compartment, a small quantity of moist food was smeared on the brass panel immediately above the lever. This tended to orient the rat's exploratory movements toward the vicinity of the lever so that, within a relatively short time, the rat would 'accidentally' depress the lever, and a pellet of food would appear in the tray. Three or four such accidental bar depressions might occur in this way, before the rat commenced to eat the food *immediately* after depressing the bar. *E* counted the first response which was immediately followed by the ingestion of the pellet as the *first reinforced response* and, thereafter, the rat was allowed to remain in the box until the impulse counter registered 70 or 90 reinforcements, as the case may be.

(4) *Satiation and extinction.*—On the twelfth day, each animal was satiated as thoroughly as possible, and immediately thereafter the strength of any remaining tendency to perform the bar-depression response was tested by extinction to a five min. no response criterion. Satiation was accomplished by placing a large container of moist food in the home cage of the animal at the customary time for experimentation. The container of food was left in the cage for a minimum time of one hour. If, at the end of about 55 min., the animal had stopped eating, *E* alternately offered it moist and dry food by hand until the food was refused. After this, the animal was

[7] When Perin's theoretical equation (8, pp. 108–111) for number of reactions to produce extinction as a function of number of reinforcement repetitions and strength of hunger is set at zero hunger, and 70 and 90 reinforcements respectively, the resulting difference is less than one full extinction reaction.

left in the cage for an additional two min. to determine whether it would resume eating from the container. Although food was available to each animal for a minimum of 60 min., in no case did it require more than 70 min. for the animal to attain this two min. criterion of food refusal.[8] Immediately after satiation, each rat was placed in the apparatus for the extinction trials. The food delivery device was disconnected, so that any depression of the bar would be followed neither by a pellet nor the characteristic thump of the food delivery mechanism.

In accordance with Perin's procedure, it was planned to measure the strength of the bar-depression behavior tendency (reaction potential) in three ways: "(1) the number of extinction (non-rewarded) responses before the 5-minute criterion of no response; (2) the time required for extinction exclusive of the 5-minute criterion; (3) the latency or time required to make the first three extinction responses, a single latency being the temporal interval between the point in time at which an animal was placed in the experimental compartment and that at which it pushed the bar" (8, p. 98). In determining the latencies of the first three extinction responses, Perin tried to minimize the effects of the previous extinction on the latencies of the second and third responses by returning the rats to their home cages for a one min. spontaneous recovery interval after the first and second responses. In order to maintain a parallel routine, the same plan was adopted in the present experiment. All animals were given three extinction trials separated by the one min. recovery interval, even in cases where a rat satisfied the extinction criterion on the first or second trial. Thus, three 'latencies' were determined for every animal. *E* timed each latency with an electric timer. After the performance of the third response, the animal was left in the apparatus until attainment of the five min. extinction criterion. Total time preceding the criterion was measured by means of a second electric timer.

It is believed that the conditions of the present experiment constitute an accurate reduplication of Perin's conditions in all essential respects. There are, however, three differences which we feel to be of distinctly minor importance. (1) In the Perin experiment the apparatus was ventilated by a suction line which sucked air out of the apparatus; in the present experiment air was blown into the apparatus. (2) Perin's pellets were made of a mixture of calves meal and water; our own consisted of ground Purina checkers, whole wheat flour and water. (3) Perin's pellets weighed .050 gm. each; our pellets weighed .075 gm. each.

It is difficult to see how the difference in the direction of the air current could prejudice our results in either direction. The difference in the composition of the pellets could influence our results only if the two types of pellets are of differential nutritive value, for Hull has assumed that the limit of the curve of habit strength is a positive function of the 'quality' of the reinforcing agent (6, p. 133). Nevertheless, even if such a difference in quality exists, it is highly doubtful that our results could be biased by this factor. If our pellets are of greater nutritive value than Perin's, this would mean that a higher habit strength value was attained during training in the present experiment than in Perin's. Such a difference could only favor a *higher* reaction potential value under zero drive than that estimated by Perin. On the other hand, if our pellets are of lower nutritive value than Perin's, this might mean that we had failed to attain a habit strength value as high as Perin's, and thus the present experiment would be loaded with a factor predisposing toward a lower reaction potential value under zero drive. This latter possibility, however, is effectively compensated for by the third difference between the conditions of the two experiments—the fact that our pellets are 50 percent heavier than Perin's. For Hull has assumed that the asymptote of the curve of habit strength is also a positive function of the *quantity* of reinforcement (6, p. 133). Thus, even if our pellets were of lower nutritive value than Perin's, the greater amount of food received per reinforcement would almost certainly eventuate in habit strength values as great or greater than those attained in the Perin experiment.

Results

The primary results of the experiment may be found in Table I. This table presents the three reaction potential measures for each of

[8] This satiation technique, like the rest of our experimental routine, is almost an exact duplication of Perin's procedure. Fortunately for the present experiment, it is an extremely effective method of satiation. A similar satiation technique has been found to be vastly superior to the method of making unlimited food continuously available to the animal, in a number of unpublished experiments conducted in the Duke laboratory by Zener and others.

the 32 rats employed in the present experiment. It is clear that the reaction potential values yielded by each rat are strikingly low. The modal reaction potential, as measured by the number of reactions preceding extinction, is actually zero, 14 of the animals having failed

TABLE I

THE THREE MEASURES OF REACTION POTENTIAL—NUMBER OF EXTINCTION RESPONSES, TIME PRECEDING EXTINCTION, AND LATENCIES OF THE FIRST THREE EXTINCTION RESPONSES—FOR EACH RAT

Number of Rat	Number of Reinforcements	Measures of Reaction Potential				
		Number of Extinction Responses	Time Preceding Extinction (Min.)	Latencies of the First Three Extinction Responses * (Min.)		
				1	2	3
1	90	1	.15	.15	—	—
2	90	1	1.18	1.18	—	—
3	90	0	0	—	—	—
4	90	0	0	—	—	—
5	90	3	2.17	.63	.62	.92
6	90	2	.83	.40	.43	—
7	70	2	1.23	.38	.85	—
8	70	2	1.87	.65	1.22	—
9	70	0	0	—	—	—
10	70	3	1.62	.30	.15	1.17
11	70	1	2.07	2.07	—	—
12	70	1	.27	.27	—	—
13	70	1	.45	.45	—	—
14	70	0	0	—	—	—
15	70	1	.73	.73	—	—
16	70	0	0	—	—	—
17	70	0	0	—	—	—
18	70	0	0	—	—	—
19	70	1	1.12	1.12	—	—
20	70	2	2.48	1.38	1.10	—
21	70	0	0	—	—	—
22	70	1	.73	.73	—	—
23	70	3	11.70	4.18	4.52	3.00
24	70	0	0	—	—	—
25	70	0	0	—	—	—
26	70	0	0	—	—	—
27	70	1	1.05	1.05	—	—
28	70	2	.95	.42	.53	—
29	70	0	0	—	—	—
30	70	0	0	—	—	—
31	70	0	0	—	—	—
32	70	1	1.58	1.58	—	—

* Whenever a dash occurs in the latency columns, this means that the rat made no response in a five min. period. Such a latency is indeterminate, having some value between five min. and infinity, because if the animal did not respond in five min., it was removed from the apparatus and returned to its home cage. The reliability of the five min. extinction criterion was checked by going through the procedure of determining three 'latencies' for all animals, even in the case of rats who satisfied the criterion on the first or second trial. Thus, an animal who did not respond for the initial five min. in the apparatus was returned to the home cage for a one min. recovery interval, and reintroduced into the apparatus for a second and third trial. In no case did an animal who satisfied the extinction criterion on the first or second trials, fail to satisfy the criterion on the subsequent trials which it received.

to make a single response before attainment of the five min. extinction criterion. The range of the resistance to extinction values is only from zero to three.

As might be expected from the low resistance to extinction values, the two remaining reaction potential measures also show very low values. The time preceding satisfaction of the extinction criterion (exclusive of the criterion) ranges from the modal value of zero to 11.70 min. (rat 23). The distribution of time scores, however, is much more homogeneous than the range seems to indicate, for rat 23 with a score of 11.70 min. deviates markedly from the trend of the distribution, the next highest score being only 2.48 min.

Continuing to the latency values, and recalling that a high latency indicates a low reaction potential, it is clear that the latencies are uniformly high, ranging from .15 min. to an indeterminate latency somewhere between five min. and infinity. The 'indeterminate' latencies correspond, of course, to cases where the rats failed to respond a single time during a five min. extinction interval. Since five min. of no response constitutes satisfaction of the extinction criterion, the rats were removed from the apparatus at the end of this interval, thus leaving the true latency at some undetermined point between five min. and infinity. Each dash in Table I represents such an indeterminate latency. Consequently, it is clear from the table that the modal values for each of the three initial responses are indeterminate. Combining the latencies for all three responses into a single distribution, 67 out of the entire population of 96 latencies turn out to be indeterminate.

Examination of Table I also discloses that the results for rats 1–6, who received 90 reinforcements during training, are substantially the same as the results for the 26 70-reinforcement animals. The distributions of the number of extinction responses for the two groups of animals have identical ranges (1–3), and identical medians (1). For this reason, it seems legitimate to treat the 70 and 90 reinforcement animals as a single group, and this procedure will be adhered to in presenting the rest of the results.

Table II presents the average reaction potential values obtained in the present experiment, and compares them with the theoretical values predicted by Perin's equations. The average values for the present experiment show, even more strikingly than the individual scores, how closely our results approximate a minimal reaction potential. The mean number of responses preceding extinction is .88 and the median is one. The mean time preceding extinction is 1.05 min., and the corresponding median is only .36 min. Since a large number of the latencies are indeterminate, the calculation of a mean latency for the three initial extinction responses becomes impossible. When the

TABLE II

THE OBTAINED AVERAGE VALUES FOR REACTION POTENTIAL AT MAXIMUM REINFORCEMENT AND ZERO HUNGER COMPARED WITH PERIN'S EXTRAPOLATED VALUES

Measure of Reaction Potential		Values Obtained in the Present Experiment Hunger = 0 Reinforcements: 26 rats = 70 6 rats = 90	Perin's Extrapolated Values *	
			Hunger = 0 Reinf. = 70	Hunger = 0 Reinf. = 90
(1) Number of Extinction Responses Preceding the Five Min. Criterion	Mean Median	.88 1.00	16.27	16.93
(2) Time Preceding Extinction (Min.)	Mean Median	1.05 .36	9.20	9.40
(3) Latency of Initial Extinction Responses (Min.)	Median	Indeterminate (5 min. — ∞)	.52	.54

* These values were derived from Perin's equations (8, p. 108–111) for the three measures of reaction potential as a function of number of reinforcements (R) and number of hours of hunger (h). These equations are:

(1) $$N = 21.45 \times 10^{.0222h}(1 - 10^{-.018R}) - 4.$$

(2) $$T = 11.0 \times 10^{.00053h^2}\left[1 - 10^{\frac{R}{16}\log\left(1-.567\times10^{-.00025h^2}\right)}\right] - 1.5.$$

(3) $$L = \frac{(.43 - .01h)R + (7.22 - .275h)}{R - 1}.$$

'N' stands for the number of extinction reactions, 'T' for time preceding extinction and 'L' for latency. Since the curves were fitted to median values throughout, all solutions of these equations for N, T, and L represent medians. For this reason, it was impossible to compute theoretical *means* for Perin. The extrapolated median values attributed to Perin in the table were arrived at by solving each of the equations after setting R at 70 and h at 0, and R at 90 and h at 0, respectively.

latencies for the first three extinction responses are combined into a single distribution, the median turns out to be itself indeterminate (i.e., somewhere between 5 min. and infinity). Although the significance of such a median is ambiguous, the 'indeterminate' score does represent the maximum latency obtainable under the conditions of the present experiment, and so may meaningfully be compared with Perin's theoretical latency values.

The differences between the average reaction potential values obtained in the present experiment and Perin's theoretical values are impressive. Table II presents values derived from Perin's equations for the three measures of reaction potential when number of hours' hunger is set at zero, and number of reinforcements at 70 and 90, respectively. Since these equations were fitted to curves plotted in terms of medians, Perin's theoretical values can only be computed in terms of medians. Values were derived for *both* 70 and 90 reinforcements in order to show the virtual identity of the reaction potential estimates at these points. When the median reaction potential values

obtained in the present experiment are compared with Perin's extrapolated values, the following marked differences emerge. (1) Perin's medians for the number of responses preceding extinction are more than 16 times as great as our obtained value. (2) The Perin figures for time preceding extinction are over 25 times as great as the median obtained in the present experiment. (3) If we assume that our 'indeterminate' median latency is the *minimum* value of 5 min., then Perin's extrapolated latency values become almost 20 times smaller than our obtained value. Since reaction potential is the reciprocal of latency, this latency difference means that Perin's theoretical reaction potential, as measured by latency, is 20 times greater than the present obtained value.

The significance of our results, relative to the Perin-Hull theoretical predictions, is shown—perhaps in a more meaningful way—in Table III. The meaning of the absolute reaction potential values may be more readily appreciated when they are presented as percentages of the values that would result under conditions of maximum reinforcement and maximum drive. Accordingly, Table III shows how the reaction potential values at *zero hunger* and maximum reinforcement (70 and 90) compare with the values that would result at *maximum hunger* (22 hours) [9] and maximum reinforcement. This comparison is made both for the zero hunger values obtained in the present experiment and Perin's zero hunger extrapolations. The estimates of reaction potential at *22 hours of hunger*, and 70 and 90 reinforcements, are derived from Perin's equations. These estimates are probably fairly accurate, because one of the groups [10] included in the data to which the curves were fitted was actually extinguished at 90 reinforcements and 22 hours of hunger.

Table III shows the impressive contrast between the ratios of reaction potential at zero hunger, relative to the theoretical value at 22 hours' hunger, obtained in the present experiment, and the ratios derived from Perin's equations. When reaction potential is measured by resistance to extinction, the present experiment yields a zero/22 hours' hunger percentage of 1.71, while Perin's theoretical percentage is 27.85. When time preceding the extinction criterion is taken as the measure of reaction potential, our obtained ratio is 1.97 percent, while Perin's extrapolated value jumps to 50.25 percent. Because of the indeterminacy of the obtained median latency value, it is impossible to compute an unambiguous zero/22 hours' hunger latency percentage

[9] Twenty-two hours is the maximum period of food deprivation for which the Perin equations are computed. Thus, when we use the terms 'maximum hunger' or 'maximum drive,' we mean the maximum in terms of Perin's experiment, not the maximum strength which might be obtained after longer periods of deprivation.

[10] This 90 reinforcement and 22 hour hunger group was one of the groups run by Williams (12).

TABLE III

COMPARISON OF REACTION POTENTIAL VALUES AT *Zero Hunger* AND MAXIMUM REINFORCEMENT WITH THE THEORETICAL REACTION POTENTIAL VALUES AT *22 Hours' Hunger* AND MAXIMUM REINFORCEMENT DERIVED FROM PERIN'S EQUATIONS

Measures of Reaction Potential	Minimum Hunger Maximum Reinforcement		Maximum Hunger Maximum Reinforcement *		Ratio of Reaction Potential at 0 Hunger and 70 Reinf. (1) to Reaction Potential at 22 Hrs.' Hunger and 70 Reinf. (2) (Expressed in Percent)	
	(1) Hunger = 0 Reinforcements = 70		(2) Hun. = 22 Hrs. Reinf. = 70	Hun. = 22 Hrs. Reinf. = 90		
	Obtained in Pres. Exper.	Perin's Extrapolation	Perin's Extrapolation	Perin's Extrapolation	Obtained Val.: Value at 22 Hrs.' Hunger	Perin's Extrap. Val.: Val. at 22 Hrs.' Hunger
Median Number of Extinction Responses Preceding the Five Min. Criterion	1.00	16.27	58.42	60.52	1.75%	27.85%
Median Time Preceding Extinction (Min.)	0.36	9.20	18.31	18.34	1.97%	50.25%
Median Latency of Initial Extinction Responses (Lat. = reciprocal of reaction potential)	Indeterm. (5 min.−)	.52	.23	.24	Indeterm. (0–10.40%)	44.25%

* We present reaction potential values at 'maximum' reinforcement and 'maximum' hunger for *both* 70 and 90 reinforcements because some of the animals in the present experimental group received 70 reinforcements, and some received 90. The closeness of Perin's theoretical values for 70 and 90 reinforcements, even at 22 hours' hunger, seems to justify: (1) our practice of allowing most of the animals only 70 reinforcements in the attempt to approximate maximum habit strength, and (2) the combination of the 26 70-reinforcement animals and the six 90-reinforcement animals into a single group. The two sets of reaction potential values at maximum reinforcement and maximum hunger which appear in the tables are derived from Perin's three equations (footnote to Table II) by setting R at 70 and h at 22, and R at 90 and h at 22, respectively. The data to which Perin fitted his equations do not contain a group which was extinguished at 70 reinforcements and 22 hours' hunger, but they do contain a group—deriving from Williams' (12) experiment—extinguished at 90 reinforcements and 22 hours' hunger. The theoretical 90 reinforcement values derived from Perin's equations correspond closely to Williams' obtained 90 reinforcement values, and so can be considered as substantially accurate.

for our results. In consequence, our obtained latency percentage may best be represented as indeterminate. However, if we arbitrarily set our median obtained latency at the minimum value of five min., the percentage of reaction potential at satiation relative to that at 22 hours' hunger becomes 10.40. Thus, the obtained 'indeterminate' ratio may be estimated as between zero and about 10.40 percent. On the other hand, Perin's theoretical latency percentage is 44.25.

A partial check on the degree to which the present experiment succeeded in reduplicating the conditions of the Perin experiment may be found in Table IV. This table compares the initial learning ability of the present experimental group with that of Perin's groups, in terms of the time required for the performance of the first five rein-

TABLE IV

COMPARISON OF THE AVERAGE LEARNING ABILITY OF THE PRESENT EXPERIMENTAL GROUP WITH THE LEARNING ABILITY OF PERIN'S GROUPS

Measure of Initial Learning Ability		Present Exp'tal Group	Perin's Groups							
			Group *A*				Group *B*			
			1	2	3	4	1	2	3	4
Time for First Five Reinforcements During Training (Min.)	Mean	2.57	2.31	2.40	2.10	2.21	2.37	2.15	2.24	2.32
	Median	2.29	2.20	2.10	1.85	1.85	2.20	1.95	2.10	2.00
	σ Mdn.	1.32*	.21	.21	.16	.21	.28	.19	.19	.21

* The very large size of our sigma, relative to Perin's, is due to the inclusion in our results of some six animals with extremely high time scores for the first five training responses. It was felt legitimate to include these animals because their performance was comparable to that of the other animals in all other respects.

forced responses during training.[11] Here it will be seen that the mean time required by the present group for the first five reinforced bar depressions is 2.57 min., while the corresponding means for Perin's eight groups range from 2.10–2.40 min. Although our mean is slightly higher than the mean of Perin's slowest group, the difference between these two values (.17 min.) is considerably smaller than the difference between Perin's slowest and fastest groups (.30 min.). A similar relation holds between our median time score and Perin's medians. The median for the present group (2.29 min.) is .9 min. higher than the median of Perin's slowest group. But Perin's scores range from 1.85 to 2.20, a range almost four times the size of the difference between the present group and Perin's slowest group. Thus, the present group seems distinctly comparable to Perin's groups with respect to initial learning ability.

DISCUSSION

It will be recalled that the purpose of our experiment is twofold: (1) to test the hypothesis that strong habits may acquire intrinsic motivational properties, and (2) to check the Perin-Hull reaction potential extrapolations at zero drive and maximum habit strength.

[11] The time required for the first five reinforced responses is the only measure of learning ability which Perin presents (**8**, p. 98) for his various groups. Thus, this is the only measure available for a comparison of the learning ability of the present group with Perin's groups.

With regard to the first objective, the present results are decidedly negative. Our procedure seems to realize all of the conditions which, according to those who impute motivational power to strong habits, should eventuate in a substantial residual behavior tendency after disappearance of the original learning motivation. Nevertheless, after establishing a habit of maximum strength, the habit mediated behavior seems almost completely to disappear when the motivation of original learning is eliminated. This outcome conflicts markedly with the implications of such well-known theses as 'the functional autonomy of motives,' 'externalization of drive,' and the 'transformation of mechanisms into drives.' On the other hand, our results support the general expectation of field theorists that, no matter how strongly learned a behavior tendency may be, the behavior tendency will not be manifest during the absence of relevant motivation (need, tension).

At this point, a question might be raised as to the significance of the very low reaction potential values emerging from the present experiment. Is our mean resistance to extinction value of .88 response, for instance, to be taken as evidence for a very small but definite tendency for the bar-depression habit to acquire intrinsic motivational momentum? A number of considerations make it more reasonable to regard our small obtained reaction potential values as indicating, within the limits of experimental error, *a true value of zero* or below. Substantial support for this interpretation is provided by the fact that almost half of the experimental animals actually yielded zero reaction potential values. Even more decisive support, however, derives from qualitative observation of the animals during the first of the three extinction periods. The typical behavior sequence at this time fell into two phases: (1) For an initial one or two min., the animals would perform active exploratory movements, sniffing the walls and floor of the experimental compartment, occasionally rearing and manipulating some part of the wall, the water bottle, etc. (2) After this brief period of activity, the animals typically became quiescent, usually sitting or lying down at the rear of the compartment until they were removed by *E*.[12] During the active phase, all animals would sooner or later perform some exploratory activity in the vicinity of the brass panel. This activity was oriented mainly toward the nozzle of the air tube which was situated in the ceiling of the compartment directly above the panel. The activity sequence typically included rearing and sniffing in the direction of the air current, manipulating the region of the wall under the air vent, scratching irregularities on the

[12] During the second and third extinction periods, the active phase tended to appear in abbreviated form, if at all. Most of the rats assumed a quiescent position in the rear of the apparatus *throughout* these later extinction trials.

surface of the panel, etc. It was in the course of this interval of random activity in the vicinity of the panel that all bar depressions preceding the extinction criterion occurred. Such reactions differed markedly in form, amplitude, and behavior context from the bar depressions which occurred during the training period. Since the bar depressions present during extinction seem to be essentially different reactions from the training depressions, it is not likely that the two types of reaction are mediated by the same habit structures. It seems probable, then, that the behavior measured during extinction had no relation to the habit of original learning, and, thus, that the true reaction potential value indicated by our results is actually zero.

The unequivocal nature of our results may be largely attributed to the effectiveness of the modified Skinner-box procedure as a technique for measuring behavior at specified degrees of drive and habit. No other standard animal learning procedure provides a situation so likely to be uncomplicated by uncontrolled stimulus and drive factors. Among the reasons for this seem to be the following. (1) The 12-in. cubical compartment surrounds the animal with a thoroughly homogeneous stimulus field, small in area and limited in possibilities for exploration. (2) At the same time, the compartment permits as much freedom of activity as does the home cage, and it is generally as well ventilated and comfortable. (3) The availability of water within the compartment holds thirst to a minimum throughout the experiment. (4) The partial sound-proofing of the apparatus rules out or minimizes most uncontrolled sounds. (5) The enclosed and sound insulated character of the experimental compartment reduces any possible influence of *E* on the animal's behavior to a minimum. The fact that previous attempts to measure the strength of a behavior tendency at zero drive and maximum habit strength were made in situations—such as the maze (**3**) and salivary conditioning (**5**, **14**)—which do not share all of the above properties, may account in part for the discrepancy between earlier results and our own. We will return to an analysis of previous experimentation in a later part of the present section.

Continuing to a consideration of our second objective, it is clear that the present results are in decided disagreement with Perin's extrapolated reaction potential values for zero drive and maximum habit strength. Since Hull has based the quantitative formulation of a key postulate in his theory of behavior—postulate 7 (**6**, p. 155)—on Perin's equations, this disagreement becomes a matter of some importance. It should be worth while, then, to consider some of the consequences of this divergence between the predicted and obtained values for Perin's equations in particular, and Hull's system in general.

It will be recalled that, according to Perin's equations, even when

drive is zero, reaction potential should be a positive growth function of habit strength (number of reinforcements), approximating an asymptote of substantial magnitude. But our obtained value for this asymptote (at zero drive and maximum reinforcement) was in the neighborhood of zero. From this it follows that reaction potential at zero drive should approximate zero at all habit strength values below the maximum. Thus, the entire range of reaction potential values at zero drive is invalidated, while a still wider range of values derivable from the equations for low intensities of drive (between zero and three hours) becomes highly suspect.

An examination of Perin's three dimensional, graphical representation of his equations (**8**, pp. 108–110) discloses that each point of the zero-drive reinforcement curve serves as the origin of a curve of reaction potential as a function of hunger at the corresponding number of reinforcements. Each one of these hunger curves shows slight positive acceleration, originating from a positive reaction potential value (at all numbers of reinforcements after five), and terminating in a much higher reaction potential value at 22 hours' hunger. Our results suggest a radical revision in the form of these curves. Assuming that the reaction potential values at higher intensities of hunger are substantially correct, each hunger curve must have its origin at a reaction potential value of zero or below, rise rapidly to some point between one and three hours of hunger, and then continue, with the original slight positive acceleration, to the 22 hour hunger level. It is obviously impossible to specify the form of the initial part of the hunger curve without an empirical determination of reaction potential at a number of points at very low intensities of hunger (between, say, zero and three hours). If Perin's reaction potential values at higher strengths of hunger are correct, this revision of the hunger curve would reveal an interesting fact about the effect of very low intensities of hunger (and, possibly, drive in general) on behavior; namely, that slight increments of hunger immediately after satiation (zero hunger) have a greater relative energizing effect on behavior than similar increments taken at any higher intensity of hunger.

Since our results force a revision of Perin's equations, they must also force a revision in the quantitative formulation of Hull's postulate 7 which, we will recall, constitutes a restatement of Perin's function. The more general *verbal* formulation of this postulate—which asserts merely that reaction potential is the product of 'an increasing function of habit strength' multiplied by 'an increasing function of drive'—may, of course, still hold, in spite of any inaccuracy of the equations at low intensities of drive. Nevertheless, the marked inaccuracy of the equations at low levels of drive is capable of mediating many false

derivations within Hull's system. A number of such deductions—apparently generated in part by solutions of the equations when set at very low values of drive—have already been made. To take one of the most important of these as an example, Hull presents as corollary VII in the chapter of *Principles of behavior* on primary motivation the following:

> Under the conditions of the satiation of the dominant drive involved in the original habit acquisition, there are sufficient residuals of other drives which in the aggregate yield on the average an excitatory potential amounting to around 20 percent of that mobilized by a 24 hour hunger on a habit originally set up on the basis of this drive (6, p. 248).

In the present experiment, however, the antecedent conditions of this corollary were fully realized, and little more than zero reaction potential was obtained.

Since it is unreasonable to suppose that, even under the conditions of the present experiment, *all* residual drives were eliminated, the fact that a near zero reaction potential was obtained tends to cast doubt on another assumption presupposed by corollary 7 and related derivations: namely, that 'irrelevant,' residual drives may 'operate in a multiplicative manner upon the habit strength effective at the moment in determining reaction potentiality' (**6**, p. 248). For this reason, two additional corollaries which are based on the assumption just stated—corollaries VIII (**6**, p. 248) and X (**6**, p. 249)—fall under suspicion. It should be emphasized, of course, that the negative implications of our results with regard to the possibility of irrelevant drives, absent during the learning of a habit, combining with that habit in the production of reaction potential, are far from conclusive. A direct experimental attack on this problem is decidedly in order.

A final word may be devoted to the disagreement between our results and earlier experimental data from which a similar measurement of behavior strength at zero drive and high habit strength may be calculated. Hull (**6**, p. 231) has cited the results of Zener and McCurdy (**14**), Finch (**5**), and Skinner (**11**) as supplying corroborative evidence for Perin's extrapolated zero drive reaction potential values. It will be remembered that, when reaction potential is measured by resistance to extinction, Perin's equations imply that the asymptote of the curve of reinforcement at zero drive is 28 percent of the asymptote at 22 hours' hunger. According to Hull, Zener and McCurdy report that ". . . the mean salivary secretion from four dogs average at satiation 24 percent as much as at 21 to 24 hours' food privation." In addition, "Finch has shown that at satiation a conditioned salivary reaction in nine dogs yielded a mean of 24 percent as much secretion as was yielded at 24 hours' food privation." Finally, "measurements of one of Skinner's published graphs . . . indicate that his animals

displayed approximately 17 percent as much food seeking activity at satiation as at 25 hours' food privation" (**6**, p. 231).

Examination of the data from which these estimates were derived reveals a number of factors which may account for the marked discrepancies from the present results. Zener and McCurdy are careful to state that *complete satiation* was not attempted, but rather, "the experimental aim was to reduce hunger as low as possible without preventing the animals from eating" (**14**, p. 326). This procedure was necessary in order to make possible the recording of unconditioned secretion. In addition, Zener and McCurdy report results for *two* experimental series differing primarily in the use of a more rigorous satiation procedure in the first series than in the second. The 24 percent (actually 25 percent) ratio of conditioned secretion in the satiated state relative to the hungry state was obtained in the second series. In the first series, when zero hunger was more closely approximated, the corresponding value was only 15 percent. But even this figure must not be interpreted at its face value. One of the main findings of Zener's experiment was a marked tendency for the rate of *spontaneous* secretion to increase in the satiated as against the hungry condition. A large, but indeterminate portion of the 15 percent value may consequently be attributed to the increased spontaneous secretion present during the partial satiation. Unfortunately, spontaneous secretion was not measured during either of the series from which the above values were taken, and thus the precise contribution of this factor to the total secretion during partial satiation cannot be estimated.

Analysis of Finch's data discloses that the 24 percent estimate which Hull derived from these results is also open to question. Finch reports data for both the conditioned and unconditioned salivary secretion occurring at various intensities of hunger. Zener and McCurdy (**14**, p. 336) have pointed out that four of Finch's six dogs show marked decreases in unconditioned secretion from satiation to the 24 hour level of hunger. On the other hand, the two remaining dogs show a marked increase. The very small average unconditioned secretions shown by these two dogs during satiation suggest that the averages included cases in which the dogs did not eat the reinforcing food. If this is the case, the two dogs may be set aside from the other four as a sub-group characterized by more complete satiation. When the average ratio of conditioned secretion at satiation relative to the conditioned secretion at 24 hours' hunger is computed for this more thoroughly satiated sub-group, a value of only 5.2 percent is found. In addition, the increase in spontaneous secretion during satiation reported by Zener and McCurdy would inflate any zero-drive reaction potential estimate based on Finch's results. It is likely that this

unrecorded factor accounts for a large part of the conditioned salivation obtained during satiation.[13]

Skinner's 17 percent value derives from a learning situation much more closely resembling that of the present experiment than does the salivary conditioning situation used by Zener and McCurdy, and Finch. Nevertheless, a number of factors may account for the difference between Skinner's results and our own. For one thing, Skinner satiated his rats by the conventional technique of making food continuously available for 24 hours. As we already indicated, this technique has proven distinctly inferior to the method used in the present experiment. Again, the strength of the bar-depression tendency was measured throughout the experiment in terms of the number of bar depressions performed in test intervals of an hour. This factor alone would eventuate in the presence of slight but definite degrees of hunger during the satiation test, particularly in its later stages. If the indications of the present experiment are correct, such slight degrees of hunger might be expected to exert a very substantial effect on the strength of the measured behavior.

Summary

The present experiment attempted to measure the strength of a behavior tendency mediated by a habit of maximum strength, when the motivation of original learning is reduced to zero. Accordingly, 32 rats were trained in a modified Skinner-box situation under conditions of strong hunger. Twenty-six of the rats were allowed 70 reinforcements of the bar-depression reaction, and 6 rats received 90 reinforcements. On the next day, the animals were satiated in their home cages, and, immediately thereafter, the bar-depression reaction was extinguished in the apparatus. This procedure made it possible to measure the strength (reaction potential) of the bar-depression tendency during satiation in three ways: (1) number of bar depressions preceding a five min. no response interval, (2) time preceding the five min. extinction criterion, and (3) latencies of the first three extinction responses.

The results indicate that, under these circumstances, the strength of the bar-depression behavior tendency is strikingly close to zero. The median number of responses preceding extinction turns out to be one; median time preceding extinction is .36 min.; and the median latency was some indeterminate value between 5 min. and infinity. Another way of expressing these results is in terms of the *ratio* of the reaction potential at zero hunger and maximum reinforcement (habit strength) to the reaction potential that would result at 22 hours'

[13] The writers are indebted to Dr. K. E. Zener for bringing these points to their attention.

hunger and maximum reinforcement. Both the data and the fitted equations presented by Perin (**8**) make possible the estimation of the 22 hour hunger reaction potential values. When the three reaction potential values obtained in the present experiment are converted into percentages of Perin's 22 hour hunger values, they become: 1.71 percent for resistance to extinction, 1.91 percent for time preceding extinction, and an indeterminate value between zero and 10.4 percent for latency. The very small magnitude of the obtained averages, and the apparent difference of the bar reactions performed during extinction from the training reactions, suggest that the true reaction potential value is actually zero.

These results are significant in two ways:

(1) They are inconsistent with hypotheses such as 'the functional autonomy of motives,' which attribute intrinsic motivational properties to strong habits. Hypotheses of this sort imply that, under conditions similar to those realized in our experiment, a substantial tendency toward the learned behavior should be in evidence after satiation of the original learning motivation. Our near zero results are in marked conflict with this expectation.

(2) The obtained reaction potential values diverge radically from the corresponding theoretical values derivable from the equations of Perin (**8**) which purport to express reaction potential as a quantitative function of habit strength and drive intensity. Although the present experiment duplicates the main aspects of Perin's procedure, his extrapolated reaction potential estimates for zero drive and maximum habit strength are many times the size of our obtained values. When represented as percentages of the reaction potential at 22 hours' drive and maximum habit strength, Perin's values are approximately 28 percent (resistance to extinction), 50 percent (time preceding extinction), and 44 percent (latency). Since Hull (**6**) has adopted Perin's equations for the quantitative formulation of his chief postulate on primary motivation—postulate 7—our results invalidate the implications of this postulate for zero drive, and possibly for weak intensities of drive in general. The present form of this postulate leads to a number of derivations in Hull's system which, in the light of our results, are probably faulty.

Two subsidiary suggestions of general interest emerge from an analysis of the present results:

(1) If our obtained reaction potential values for zero hunger and Perin's values for medium and high degrees of hunger are substantially correct, then slight increases in hunger immediately after satiation seem to exert a greater relative energizing effect on behavior than similar increases at all higher strengths of hunger.

(2) The fact that we obtained near zero reaction potential values

at zero hunger in spite of the probable presence of uncontrolled drives, makes questionable the assumption of any marked tendency for *irrelevant* drives to combine with habit structures in the production of behavior.

(Manuscript received November 8, 1944)

Bibliography

1. Allport, G. W. *Personality.* New York: Holt, 1937.
2. Anderson, E. E. The externalization of drive: I. Theoretical considerations. *Psychol. Rev.*, 1941, **48**, 204–224.
3. Anderson, E. E. The externalization of drive: III. Maze learning by non-rewarded and by satiated rats. *J. genet. Psychol.*, 1941, **59**, 397–426.
4. Ellson, D. G. Quantitative studies of the interaction of simple habits: I. Recovery from specific and generalized effects of extinction. *J. exp. Psychol.*, 1938, **23**, 339–358.
5. Finch, G. Hunger as a determinant of conditional and unconditional salivary response magnitude. *Amer. J. Physiol.*, 1938, **123**, 379–382.
6. Hull, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.
7. Lewin, K. The conceptual representation and the measurement of psychological forces. *Contr. psychol. Theor.*, 1938, **1**, No. 4.
8. Perin, C. T. Behavior potentiality as a joint function of the amount of training and the degree of hunger at the time of extinction. *J. exp. Psychol.*, 1942, **30**, 93–113.
9. Perin, C. T. The effect of delayed reinforcement upon the differentiation of bar responses in white rats. *J. exp. Psychol.*, 1943, **32**, 95–109.
10. Siegel, P. A note on the construction of Skinner box pellets. *J. comp. Psychol.*, 1944, **37**, 233–235.
11. Skinner, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
12. Williams, S. B. Resistance to extinction as a function of the number of reinforcements. *J. exp. Psychol.*, 1938, **23**, 506–521.
13. Woodworth, R. S. *Dynamic psychology.* New York: Columbia University Press, 1918.
14. Zener, K. E., & McCurdy, H. G. Analysis of motivational factors in conditioned behavior: I. The differential effect of changes in hunger upon conditioned, unconditioned, and spontaneous salivary secretion. *J. Psychol.*, 1939, **8**, 321–350.

DRIVE INTERACTION: II. EXPERIMENTAL ANALYSIS OF THE ROLE OF DRIVE IN LEARNING THEORY

BY HOWARD H. KENDLER[1]

State University of Iowa

On the basis of experimental data involving the use of the classical (Pavlov) and instrumental (Skinner) conditioning situations, Hull (3) has proposed a system of interrelated hypothetical variables, the purpose of which is to integrate learning data. As Hull conceives such constructs, they serve as intervening variables which help to ascertain the nature of the functional relationships (empirical laws) holding between the dependent response variables and the independently manipulable experimental (environmental) variables.

Beginning with the observed response variables Hull has proposed the following set of postulates:[2]

$$(1) \quad R = f\, {}_{S}\mathbf{E}_{R}$$

$$(2) \quad {}_{S}\mathbf{E}_{R} = f({}_{S}\mathbf{H}_{R}) \times f(\mathbf{D})$$

$$(3) \quad {}_{S}\mathbf{H}_{R} = f(N, Ts, Tg, Mg)$$

$$(4) \quad \mathbf{D} = f(Td)$$

where R = response measure, e.g., number of trials to extinction, amplitude of response, etc.;
${}_{S}\mathbf{E}_{R}$ = excitatory potential, a theoretical construct defined in terms of ${}_{S}\mathbf{H}_{R}$ and D;
${}_{S}\mathbf{H}_{R}$ = habit strength; a theoretical construct defined as a certain function of four experimental variables;
N = number of reinforcements following the response to be learned;
Ts = the time the conditioned stimulus is acting before the occurrence of the response to be learned;
Tg = time of delay of attainment of goal object following response to be learned;
Mg = magnitude of the goal object.

[1] This experiment is part of a dissertation presented to the faculty of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Professor Kenneth W. Spence for many valuable suggestions and criticisms.

[2] The bold-face symbols are hypothetical constructs or intervening variables. In Hull's system they are all ultimately defined in terms of manipulable experimental variables. See Spence (6) for an evaluation of this type of theory as opposed to other present day formulations.

D = drive strength, a theoretical construct defined as a function of *Td*;
Td = time of deprivation of some goal object or goal objects.[3]

In order to enlighten the reader as to the function of the preceding series of equations, an analysis of one of the experiments reported in the initial paper (4) on drive interaction will be made.

In an attempt to obtain information as to the nature of the empirical law existing between the rate of learning and the strength of an interacting drive, an experiment involving the use of the Skinner box was conducted (4). Five groups of white rats under different motivational conditions received 30 food reinforcements of the bar-depressing response. The hunger drive for all five groups was based upon 22 hours of food deprivation, but the thirst was varied from group to group (0, 3, 6, 12 and 22 hours of water deprivation). Following the acquisition trials the five groups were extinguished under the same motivational conditions as obtained in their conditioning trials. The results revealed that the effect of increasing the thirst drive was to increase the resistance of the response to extinction up to the value of 12 hours of water deprivation. At the next experimental value, 22 hours of water deprivation, the extinction measure was much lower, being less than that of zero hours of thirst.

Hull's constructs would be employed as follows. As the result of the training trials under the experimental conditions of magnitude and delay of reward a certain amount of habit strength to bar depressing would be developed. Under the particular drive conditions this would produce a certain excitatory or behavior potential to make the response of depressing the lever. The habit strength ($_sH_R$) and drive strength (D) are assumed to combine in a multiplicative manner to determine the behavior potential. The greater the behavior potential developed as the result of this training the greater will be the number of non-reinforced responses to reach the criterion of extinction.

It should be noted that the differential number of trials to extinguish the response for the different groups implies that different amounts of behavior potential were available at the time of the extinction series. As Hull's theory is now formulated the differences in the experimental results must be attributed to the difference in drive strength (D) of the five groups because the variables which determine the value for habit strength (see equation 3) were the same for all experimental groups. However it is conceivable that the group with 12 hours of water deprivation required the greatest number of extinc-

[3] Certain other factors (e.g., inhibition, $_SI_R$) enter into this theoretical framework, but for simplicity's sake they are omitted from the discussion.

tion trials because it developed the most habit strength during training. If this were found true, it would require that equation 3 be changed to include D as one of the variables determining the rate of development of sH_R.

The present experimental question is concerned with the adequacy of Hull's assignment of the role of drive in the formulation of his learning theory, i.e., whether the differential results obtained under the varying intensities of the additional drive in the above described experiment (4) were entirely due to the differences in drive strength at the time of testing or whether they were due to differences in habit strength or 'amount of learning' which took place during the training trials which preceded the test.

Experiment I

Experimental procedure

For the present experiment, the data of two groups (T-12 and T-22) of the previous experiment (4) were utilized. These two groups received 30 reinforcements in a Skinner box under 22 hours of food deprivation; but one of the groups (T-12) [4] was deprived of water for 12 hours while the other group (T-22) was deprived of water for 22 hours. Both groups were extinguished under similar motivational states.

The methodology of this investigation is identical with that of the previous experiment (4) except for one variation. Two groups of 20 animals, of approximately equal mean weight and age as the other two groups, were trained under the motivation of T-12 and T-22, but their drive states were reversed during extinction. If we utilize the results of the constant Groups T-12 and T-22 of the preceding experiment we can arrange a factorial type [5] of experimental design (5). This design is indicated in a two-by-two table (see Table I) with the variables being the two different degrees of thirst (T-12 and T-22) at the two different stages of the experimental procedure, acquisition and extinction.

TABLE I

Design of the Experiment

		Motivation during Extinction	
		T-12	T-22
Motivation during Acquisition	T-12	T-12(12)	T-12(22)
	T-22	T-22(12)	T-22(22)

With such a methodology we would have four experimental groups, two of which would have the identical motivational states during acquisition and extinction, while the other two would reverse their drive states. The four groups will be represented in the following manner with the thirst drive state during extinction bracketed; T-12(12), T-22(22), T-12(22), T-22(12). By an analysis of variance we may evaluate the results in terms of the drive states during acquisition and extinction, and thereby ascertain the specific locus of the influence of the additional drive.

[4] Since for all groups the degree of food deprivation was always 22 hours, reference to the experimental groups is always in terms of the degree of thirst drive. The initial experiment (4) reports in detail the specific experimental manipulations.

[5] A factorial design permits comparisons of two or more variables at once. Reference is made to Lindquist (5) for a description of this type of experimental design which can be evaluated with an analysis of variance type of statistical treatment.

Results and Discussion

The extinction scores were transformed into $1000 \times \log_{10}$ in order to fulfill the assumptions of our proposed statistical treatment.[6] The variances of the transformed scores of Groups T-12(22) and T-22(12) were 29,951 and 43,869 respectively. The two most divergent variances of this factorial design were those of T-22(22) (64,918) and T-12(22). When these variances are compared by means of an *F*-test they are significant at the 5 percent level. However, such a difference is not very great if we consider that there are six possible comparisons. Lindquist (**5**) has shown that such fluctuations do not appreciably affect the results of the analysis of variance.

Table II presents the means of the transformed scores.

TABLE II

Mean Number of Responses ($1000 \times \log_{10}$) to Reach the Criterion of Extinction

		Motivation during Extinction	
		T-12	T-22
Motivation during Acquisition	T-12	1,448	1,262
	T-22	1,359	1,209

The analysis of the simple factorial design results in four estimates of the population variance as revealed in Table III.

TABLE III

Results of the Analysis of Variance into Four Components

Component	Degrees of Freedom	Sum of Squares	Variance
Drive state during acquisition	1	83,583	83,583
Drive state during extinction	1	604,693	604,693
Interaction	1	185,555	185,555
Within cells (error)	86	4,335,491	50,413

The variance ratios in which we are particularly interested are as follows (the numbers in parentheses after each *F* being the values needed for significance at the 5 percent and 1 percent levels, respectively, for the given degrees of freedom):

Acquisition Drive/Within Cells $F = 1.66(3.96; 6.96)$
Extinction Drive/Within Cells $F = 11.98(3.96; 6.96)$

We note that only the variance for extinction drive is significant when tested against within cells. In fact it is significant at the 0.1 percent level. This reveals that the differences obtained between the

[6] This transformation resulted in an approximation of the assumptions of normality and equal variance so that the analysis of variance technique might be used to evaluate the results. This expedient redefinition of our response variable has considerable precedent in scientific reports.

groups having 12 hours of thirst as compared with the groups having 22 hours of thirst during extinction cannot be reasonably attributed entirely to fluctuations in random sampling within the experiment. The variance ratio involving acquisition drive and within cells, on the other hand, is not significant and we cannot reasonably reject the hypothesis that the motivational state during acquisition had no differential effect upon our measure of learning. These results indicate that the differential results obtained in the second experiment of the previous paper (4) can be attributed to the different motivational states at the time of extinction and not during acquisition. This evidence would suggest that (variation in) drive strength (D), at this level at least, does not determine the rate of development of habit strength ($_sH_R$).

Using Festinger's method (1) of comparing means from highly skewed samples, we obtain essentially the same results with an analysis of the raw scores. Table IV presents the means of the raw scores for the four groups which comprise the factorial design.

TABLE IV

MEAN NUMBER OF RESPONSES TO REACH THE CRITERION OF EXTINCTION

		Motivation during Extinction	
		T-12	T-22
Motivation during Acquisition	T-12	32.92	19.75
	T-22	25.30	18.88

Table V indicates the results of the comparisons between those groups which had the same motivation during acquisition but different motivation during extinction.

TABLE V

COMPARISON OF DIFFERENCES BETWEEN MEAN NUMBER OF RESPONSES TO REACH CRITERION OF EXTINCTION *

Groups	Ratio of Means	F-value	Degrees of Freedom	Level of Significance
T-12(12)/T-12(22).....	32.92/19.75	1.67	111/212	1%
T-22(12)/T-22(22).....	25.30/18.88	1.34	209/167	5%

* Comparisons are between groups that had the same motivation during acquisition but different motivation during extinction.

With the chi-square test of significance of combined probabilities we find the probabilities of Table V to be significant beyond the one percent level of confidence. Therefore, with the results of this analysis we can safely reject the null hypothesis that there is no difference between the effects of a 12 and 22 hour co-existent thirst drive during the *extinction* of the bar-depressing habit.

If we now make comparisons between groups that had the same motivation during extinction but different drive states during acquisition, we may draw conclusions as to the effect of the additional drive during acquisition on the learning scores used in this experiment.

TABLE VI

COMPARISON OF DIFFERENCES BETWEEN MEAN NUMBER OF RESPONSES TO REACH CRITERION OF EXTINCTION *

Groups	Ratio of Means	*F*-value	Degrees of Freedom	Level of Significance
T-12(12)/T-22(12).....	32.92/25.30	1.30	111/209	between 20% and 5%
T-12(22)/T-22(22).....	19.75/18.88	1.05	212/167	greater than 20%

* Comparisons are between groups that had different motivation during acquisition but the same motivation during extinction.

Neither of these differences is significant beyond the 5 percent level, although the first difference barely misses being significant at this level of confidence. The significance of the second comparison is appreciably greater than 20 percent. If, by the chi-square method, we combine 20 percent and 5 percent the resulting probability is not significant at the 5 percent level of confidence. Obviously the two obtained probabilities are greater than 20 percent and 5 percent, so the resultant probability would also not be significant at the 5 percent level of confidence. Therefore, from the results of this analysis we cannot reasonably reject the null hypothesis, namely that the different drive states at the time of acquisition have no effect upon the extinction scores. Such a conclusion is consistent with the one arrived at by means of the analysis of variance of the transformed scores.[7]

The difference between the two groups having 12 hours of thirst during extinction (T-12(12) and T-22(12)) is almost significant at the 5 percent level as is indicated in Table VI. This suggests the presence of some generalization effect when the motivational state is changed, i.e., the animals that were conditioned under 22 hours of thirst and extinguished under 12 required less trials to reach the criterion of extinction because the response being extinguished was a generalized one. However, the results of the other two groups do not support this line of reasoning since the 'generalized' 22 hour thirst group required more trials to reach extinction (see Table VI).

[7] Additional experimental evidence which support these conclusions will be presented in the latter part of the paper.

Experiment II

Introduction

In the previous paper (4) of this series of experimental articles on drive interaction, it was discovered that animals motivated by 22 hours of hunger (Group H) learned more efficiently (committed less errors) to make a simple spatial discrimination than a group motivated simultaneously by 22 hours of hunger and thirst (Group H & T).

The specific experimental manipulations were as follows. Two groups of 16 animals each received three trials per day in a T-maze, one side of which contained food while the other side was empty. The criterion of learning was 12 successive trials to the side which contained food. Group H had a mean of 2.31 errors while Group H & T had a mean of 3.69 errors.

According to our analysis of the results of Experiment I the above discrepancy between the learning scores of the two groups of the T-maze study (4) can be attributed to the stronger drive in the group motivated solely by hunger. Or, in terms of Hull's concepts, a larger value of D, one of the two components determining response strength ($_sE_R$), is present in Group H. Now if one assumes, as Hull does, that the difference in response strength of the competing responses at the attainment of the learning criterion is the same for both groups, then it follows that the habit strength for the H & T group must be greater than that of the H group at the completion of learning. Consistent with this implication is the fact that Group H & T required a greater number of reinforcements than Group H (means of 9.8 to 7.8). Number of reinforcements is one of the experimental variables of which habit strength is postulated to be a function. The greater number of reinforcements in Group H & T should thus provide a greater amount of habit strength for this group.

Further confirmation of this analysis are the results of certain further experiments that were carried out with these Ss. At the completion of their original spatial discrimination both groups were trained in the same maze under a thirst drive to take the opposite path. The goal box which was empty during the original learning now contained water.[8]

Subsequent to this the animals were once again required to learn to go to the side on which they originally found food. The purpose of this experiment was to ascertain whether the Ss which had originally been trained under the H & T condition would now learn more quickly

[8] This experimental variation was designed to test a problem which is not the immediate concern of the present report. However, it is necessary to point out that there was no significant difference in the results obtained. The mean number of reinforcements to reach the criterion of learning was approximately the same for the two groups. Group H & T required an average of 5.56 reinforcements, while Group H required an average of 5.38.

under equalized drive conditions than the animals of the H group. The implications of Hull's theory that habit strength would be greater for these Ss (H & T) would demand such experimental results.

Procedure

The equating of the drives was accomplished by splitting the two original groups into halves, one half having the motivation hunger and thirst, the other half having the motivation of hunger only. We have, therefore, four groups (eight animals in each) which can be identified by the following combinations; H(H), H(H & T), H & T(H), and H & T(H & T). The letters outside the brackets refer to the animals' motivation during the original learning, while the letters within the brackets indicate the drive state for this experiment.

Results and Discussion

On the basis of our theoretical analysis we would expect Group H & T(H) to exceed Group H(H) in relearning the initial problem for the existing drive strengths of both are equal, but Group H & T(H) presumably has a higher habit strength. For the same reason we would expect Group H & T(H & T) to reach the criterion of learning more quickly than Group H(H & T).

The results in terms of errors and their standard deviations are indicated for all four groups in Table VII.

TABLE VII

Mean Number and Standard Deviation of Errors to Reach the Criterion of Relearning

Group	Mean	σ
H(H)	4.375	3.29
H & T(H)	1.750	1.28
H(H & T)	4.625	3.29
H & T(H & T)	3.875	1.46

If we analyze our results in terms of comparisons of the groups with identical drive states for this experiment, we obtain the following differences according to the technique devised by Festinger (1):

1. The ratio between the means of Groups H(H) and H & T(H) gives an *F*-value of 2.5, which is significant (d.f. = 28 and 30) beyond the one percent level of confidence.
2. The ratio between the means of Groups H(H & T) and H & T (H & T) gives an *F*-value of 1.19, which is not significant (d.f. = 32 and 113) at the 20 percent level of confidence.

If we consider the above two comparisons to be independent estimates of the superiority of Group H & T over Group H (reference is to motivation of original learning) then we could use the chi-square test of combining probabilities from independent tests of significance (5). However the *F*-table only contains points of significance below

the 20 percent level. In spite of this, certain general inferences may be made concerning the combined results. If we combine, by the chi-square method, any probability up to 29 percent with a one percent probability the combined results will be significant at the two percent level or beyond it. In a like manner, a combination of a one percent probability with any probability between 30 percent and 86 percent will be significant at the five percent level. In the case of our second comparison an F of 1.26 would have been significant at the 20 percent level. Because of this, it seems certain that our combined results are at least significant at the five percent level. Therefore our theoretical analysis is very likely confirmed, i.e., the H & T group, because of their weaker drive strength in Experiment I, presumably acquired a greater habit strength to reach the criterion of learning to go to the side containing food.

In order to check further our previous finding concerning the superiority of a 22-hour hunger drive over a 22-hour simultaneous hunger and thirst motivation the following comparisons were made between those groups which had the same motivation in the original learning but different drive states in the present experiment.

1. The ratio between the means of Group H & T(H & T) and H & T(H) gives an F-value of 2.21, which is significant (d.f. = 113 and 30) beyond the one percent level of confidence.
2. The ratio between the means of Groups H(H & T) and H(H) gives an F-value of 1.06, which is not significant (d.f. 32 and 30) at the 20 percent level of confidence.

According to the above analysis it is likely that these combined results are significant at the five percent level. The results confirm our previous conclusion (4) that the presence of a strong additional thirst drive retards performance.

The experiments reported in this paper both support Hull's omission of drive (D) as one of the variables determining the growth of habit strength (${}_sH_R$). In the first experiment groups of animals learning under different drive strengths were equated in terms of motivation during the measurement of learning (extinction trials). It was found that equal amounts of habit growth occurred, independent of the drive conditions existing at the time of learning.

In the second experiment groups of rats under different motivational conditions reached the same amount of excitatory potential (${}_sE_R$) defined in terms of 12 successive correct responses. Since the group with the additional intense thirst drive required more trials to reach this criterion it was concluded that the additional intense thirst drive reduced the value of D (see equation 2). However, since both groups had equal values of ${}_sE_R$, the group (H & T) with the extra

drive must have acquired a greater value of ${}_{S}H_{R}$. In relearning the original discrimination, under equated drive strengths, it was discovered that the animals which originally acquired the discrimination under the drive of both hunger and thirst relearned more quickly. Therefore, Hull's formulation was again confirmed.

It should be noted that the present experiments are not concerned with the case in which the rewarded drive is directly varied. The hunger (rewarded) drive is indirectly affected in the present experiments by manipulating the intake of water. The results indicate that the variation of the additional drive does not affect the acquisition of habit strength by the stimulus situation to elicit the bar-pressing response. However, it should be emphasized that there are certain limitations to the amount of variation of the drive state which will not affect the habit strength. If the additional thirst drive is of such an extreme intensity that it actually interferes with the consumption of the food, the development of habit strength under such conditions would be affected. Presumably in such instance there would be failure of reinforcement because of failure to consume the goal object.

An experiment employing the standard Skinner-box technique has been reported by Finan (**2**) in which he attempted to discover the effect of hunger-drive variation at the time of acquisition of the response. Four groups of albino rats, deprived of food for 1, 12, 24 and 48 hours, respectively, were trained under similar experimental conditions to depress a lever to obtain food. Following acquisition all groups were extinguished two days later under the drive condition of 24 hours of privation of food. The present author analyzed Finan's results with the analysis of variance technique and the Festinger method. Both treatments indicated that the motivation at the time of acquisition resulted in significantly different amounts of resistance to experimental extinction for the four groups. Further experimentation, particularly with the simultaneous variation of drive and magnitude of reward, may explain the discrepancies which exist between our results and those of Finan.

Summary

The present experiments were designed to ascertain whether drive strength (D) was one of the variables determining habit strength (${}_{S}H_{R}$) in Hull's theoretical formulations. Different degrees of drive strength were obtained by varying the intensity of a co-existent non-rewarded thirst drive.

In the first experiment two groups of animals learned a bar-depressing response under different drive strengths. During the measure-

ment of learning (extinction trials) the animals were equated in terms of motivation. The results revealed that the different drive conditions at the time of learning did not influence the amount of habit growth.

In the second experiment two groups of rats under different drive strengths reached the same criterion of 12 successive correct responses in a T-maze situation. The animals with the lower drive strength required more trials to attain this level of proficiency. In the relearning of the original spatial discrimination the drive strengths of the two groups of animals were equated. The group which, during the original learning had a lower drive strength, relearned more quickly. Presumably this was due to the greater number of reinforcements they experienced during the original learning; number of reinforcements being one of the variables determining habit growth.

Both experimental findings are consistent with Hull's formulation and support his omission of drive strength (D) as one of the variables determining habit growth ($_sH_R$).

(Manuscript received August 14, 1944)

References

1. Festinger, L. A statistical test for means of samples from skew populations. *Psychometrika*, 1943, **8**, 205–210.
2. Finan, J. L. Quantitative studies in motivation: I. Strength of conditioning in rats under varying degrees of hunger. *J. comp. Psychol.*, 1940, **29**, 119–134.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton Century, 1943.
4. Kendler, H. H. Drive interaction: I. Learning as a function of the simultaneous presence of the hunger and thirst drives. *J. exp. Psychol.*, 1945, **35**, 96–109.
5. Lindquist, E. F. *Statistical analysis in educational research.* New York: Houghton Mifflin Co., 1940.
6. Spence, K. W. The nature of theory construction in contemporary psychology. *Psychol. Rev.*, 1944, **51**, 47–68.

THE ROLE OF ABSOLUTE INITIAL RESPONSE STRENGTH IN SIMPLE TRIAL-AND-ERROR LEARNING

BY PAUL S. SIEGEL

University of Richmond

In simple trial-and-error learning, that behavioral adjustment which affords goal attainment has been traditionally called the 'correct' response and all other situational adjustments have been termed 'incorrect.' Trial-and-error learning is marked then by the more or less gradual disappearance of the 'incorrect' responses, the 'correct' response achieving a final dominance.

Recent theoretical analysis (**1**, **2**) has regarded this kind of learning as a competition between simultaneously activated reaction tendencies. Occurrence of the 'incorrect' response is followed by a diminution in reaction tendency strength. Occurrence of the 'correct' response is followed by reinforcement and an attendant increment in reaction tendency strength. When the imbalance of strength significantly favors the 'correct' reaction tendency, this response prevails on every occasion. It occurs with an immediacy and to the exclusion of all other responses. This marks the final stage of trial-and-error learning.

Hull (**3**) has experimentally demonstrated the important role played by *relative* strength as it prevails among competing reaction tendencies at the outset of learning. This paper proposes to answer a somewhat different question.

If at the beginning of the trial-and-error process, *equality* of strength prevails among competing reaction tendencies, how is the course of learning affected by the *absolute* level of strength?

Restated, how is the course of acquisition of the 'correct' response affected by the *absolute* initial level of strength that obtains among *equally strong* competing reaction tendencies?

Let us suppose that in learning situation *W*, the competing reaction tendencies are initially equal in strength but *weak*. Let us further suppose that in learning situation *S*, the competing reaction tendencies are initially equal but *strong*. What difference emerges between the two learning curves?

METHOD

Subjects.—The experiment employed 64 albino rats, with ages ranging from 60–85 days at the beginning of the experiment. This population was drawn at random from the departmental

colonies of Duke University and the University of North Carolina. In the past several years considerable inter-colony breeding has been permitted between these two laboratories.

Group W was comprised of 32 animals, 18 females and 14 males. Group S duplicated this distribution with respect to age, sex, weight, and number.

Daily rations consisted of Purina checkers with an occasional supplement of lettuce. Water was available in the home cages at all times.

Apparatus.—A conventional T-maze was employed with animal-operated push-up type retrace prevention doors. The two doors were located one at the beginning of each arm of the T. In operation they required an upward pressure of approximately 10 gm. The start box door (guillotine type) was manipulated by *E* by means of a short length of twine. Illumination was afforded by a 150 watt bulb suspended three feet above the center of the maze. The maze measured three feet along the arms of the T (overall) and two feet along the approach alley (including the start box). The alleys were three in. in width.

The maze was situated in a quiet part of the building and a noisy electric fan served as a sound screen at all times. Runs were conducted only during the evening hours. In the free-choice series and the learning series (see procedure, stages 3 and 4), the reward mixture was liberally scattered beneath both compartments to control odor cues.

Procedure.[1]—Four convenient stages may be designated.

1. Orientation Period: All animals were given two weeks' experience in a straight alley situation. During this time they learned to manipulate retrace doors. They were run in a hungry condition and rewarded on crushed Purina.

2. Preliminary Training: By this time a temporal feeding rhythm had been well established. Throughout this stage and those that followed, both groups were run under conditions of 20 hours food deprivation (plus or minus ½ hour). A 10-sec. reinforcement interval was uniformly observed, and the reward mixture, presented in a small round tin pan, consisted of pulverized Purina (dry).

TABLE I

RUNNING SCHEDULES OF GROUP W AND GROUP S

Group W

Night	Reinforced Side		Number of Runs
1	L		1
2	R		1
3	R		1
4	L		1
		Totals	2 L
			2 R

Group S

Night	Reinforced Side		Number of Runs
1	L		4
2	R		4
3	R		6
4	L		6
5	R		6
6	R		6
7	L		6
8	L		6
9	R		6
10	L		6
11	L		6
12	R		6
13	L		6
14	R		6
		Totals	40 L
			40 R

[1] It is convenient in exposition to treat both groups together, indicating any differences in treatment. In actuality the program for Group W was completed before Group S was started.

The 10-sec. interval marked confinement in the reward compartment. After the initial trial all animals fully employed this time in eating.

During this stage, the two groups were treated differently. Group W received four forced runs in the T. Group S received 80 forced runs in the T. Forcing was effected by the addition of a stop or block at the region of choice. This served to seal off one or the other compartment while permitting entry of the opposite arm. The two running schedules are presented in Table I. This stage was introduced for the purpose of establishing in Group W the two equal but weak reaction tendencies; in Group S, the two equal but strong reaction tendencies. The selection of two reinforcements and 40 reinforcements merely represents a choice of two extremes of the strength continuum. The random order of successive runs was designed to obviate recency factors, i.e., again, to establish equally strong competing tendencies in each group.

3. Free-choice Series: To furnish a baseline for the final stage, both groups were given six free-choice trials during one nightly session. Both compartments of the T were accessible and the choice of either turn was followed by reinforcement. The trials were given in succession. When the animal had made a choice and had been permitted to feed for 10 sec., he was immediately returned to the start box for another trial. Caution was exercised throughout to see that the animal had ceased chewing before his introduction to the start box. He was then placed in the box and when he had oriented in the direction of the door, it was raised quietly by *E*. In this manner, all animals were given six trials during this nightly series.

Running times as well as choices were recorded. The running time was defined as the interval that elapsed between the raising of the start box door by *E* and the closing of a retrace door behind the animal.

Essential to the experimental design was an equality of strength between the two reaction tendencies *within* both groups. For this reason, those animals that yielded a 6–0 or a 5–1 distribution of choices were discarded. That is, a 6–0 or a 5–1 distribution was regarded as indicative of a strength imbalance in favor of the more frequently occurring response. Further, this imbalance would be exaggerated by the differential reinforcement that occurred during this series.[2] Only those animals yielding a 4–2 or a 3–3 distribution were retained. Twenty-six animals of Group W and 21 of Group S met this criterion during the free-choice series. Of the 26 in Group W, 15 yielded 3–3 distributions and 11 yielded 4–2 distributions. In Group S, there were 9 animals giving 3–3 distributions and 12 giving 4–2 distributions.

The average running time for all six trials was computed for each *S* in these two selected populations. For the 26 animals of Group W, the mean of these individual averages was 21 sec. The mean of the 21 animals of Group S was 8 sec. The difference between these two means is 13 sec. and the *t* of this difference is 2.77. With 45 degrees of freedom, a *t* of 2.69 is significant at the one-percent level.

The selection of the two populations assures approximate equality of strength between the two reaction tendencies *within* each group. A difference *between* groups in absolute level of reaction tendency strength is indicated by the difference in running times. This assumes, of course, that running time is inversely related to the strength of a reaction tendency (a commonly accepted operational definition). This fulfilled the two major demands of the experimental design. The two selected groups were then introduced to the experiment proper, the trial-and-error learning stage.

Learning Series: During the learning series, both compartments of the T were accessible, but the choice of only one was followed by reinforcement. When the animal took the opposite turn, he was confined in the empty compartment for a period of 10 sec. (non-reward).

To obviate idiosyncracies possible to the setup, a special control was introduced during this stage. Approximately one-half of Group W was reinforced for right turns and approximately one-half for left turns, i.e., the correct choice was made right for one-half and left for the other. These two sub-groups were designated on the basis of the distribution of choices yielded during the free-choice series. Those animals that gave a 3–3 distribution were arbitrarily assigned half right and half left. Those animals that yielded 4–2 distributions were divided between the right and left sub-groups in such a way that the to-be-correct response averaged for the *entire* group ap-

[2] Greater methodological purity might have been attained by employing an extinction series at this stage, i.e., no reinforcement for either choice. This would have yielded the desired information about the strength of the two responses without introducing further reinforcement complications. The weakness of the two reaction tendencies in Group W seemed, a priori, to rule this out. It is believed that the method here employed was a satisfactory substitute.

proximately .50, i.e., a 50 percent or chance level of occurrence of the *correct* response was established at the outset. For example, if an animal making four left turns and two right turns during the free-choice series was placed in the to-be-reinforced on the right subgroup, there was also included an animal that had made four right choices and two left choices. In this way, the frequency of occurrence of the correct response was made to originate at a 50 percent level. This point, representing the results of the free-choice series, is referred to in the results (Fig. 1) as Session 0.

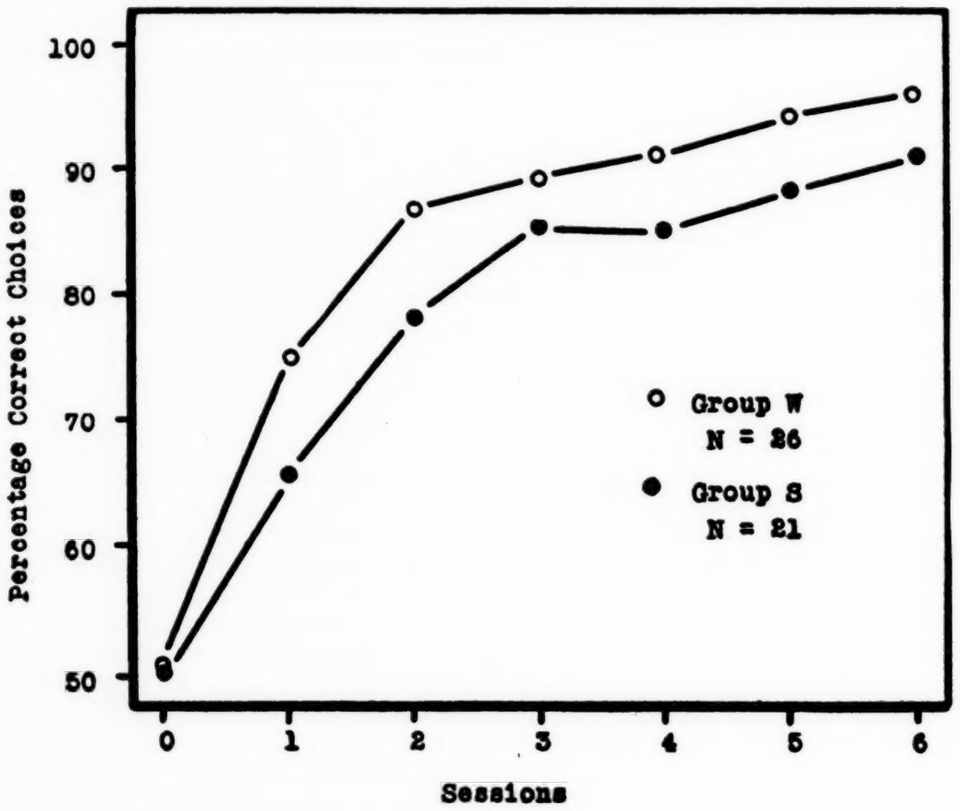

FIG. 1. Mean percentages of correct choices at each session.

Group S was similarly treated.

Both groups were then given six nightly sessions of eight trials each or 48 massed learning trials. The trials were again given in succession. The conditions of motivation and reinforcement have been indicated earlier (stage 2).

RESULTS

Our interest centers upon the difference in learning rate between the two groups, i.e., in the speed of acquisition of the correct response during the trial-and-error learning series. This interest is best served by a comparative analysis of the two curves that express the frequency of occurrence of the correct response as a function of number of trials.

Fig. 1 depicts the growth of the correct response within each group. The points on the two curves represent simple means. Each point is the mean percentage of occurrence of the correct response for the group during a nightly session of eight trials. The initial point (Session 0) has been discussed earlier. The analysis of that point is in

terms of the to-be-correct response since both responses were reinforced during the series.

Discussion

Inspection of Fig. 1 reveals a marked difference in the speed of learning of the two groups. The curve of Group W is seen to rise more rapidly. Every point (beyond Session 0, the baseline) exceeds the corresponding point for Group S.

An evaluation of the overall significance of this difference was made. For the 48 critical learning trials, the average number of correct responses was computed for each group. For Group W this mean was 43, sigma, 2.5; for Group S, 39, sigma, 4.1. The difference in means is 4 and the t of the difference is 3.96. Again, with 45 degrees of freedom, a t of 2.69 is significant at the one-percent level. We conclude that Group W responded with a significantly larger total number of correct choices.

These results rather strongly suggest that the speed of acquisition of the correct response in a simple trial-and-error situation is some inverse function of the initial level of *absolute* strength that prevails among equally strong competing reaction tendencies.

Implications for Behavior Theory

The S-R theory of Hull.—Hull (**5**) has indicated in some detail an a priori expectation of the results obtained here. Although this derivation from his latest postulate set (**4**) is far too lengthy for inclusion here, a statement of the final corollary follows:

> "In distributed-trials simple trial-and-error learning, where $_sE_R+$ and $_sE_R-$ begin with equal strength, the rate and amount of rise in the probability of reaction evocation (p) decreases and the number of trials (N) required to give $R+$ dominance increases, as the magnitude of the initial $_sE_R$'s increases (**5**, p. 30)."

$_sE_R$ is equivalent to what has been herein termed a reaction tendency, p corresponds to our percentage of correct responses, and $R+$ refers to the correct response.

Whereas this corollary was formulated for distributed learning, a later corollary makes it equally applicable to the massed learning employed in this experiment.

Clearly this anticipates the faster learning of Group W, and we must regard these results as an empirical verification with an attendant strengthening of Hull's formal postulate set.

Field theory.—It is difficult to trace the implications of our results for the systems of Tolman and Lewin. A certain embarassment is implied however for the position of White (**6**), who has attempted an interpretative analysis of Tolman and Lewin. He has explicitly done

what has been strongly implied by both Tolman and Lewin. He has subsumed the principle of reinforcement under the field principle of perceptual learning.

"In other words, there are not two sorts of learning, 'reinforcement' and 'latent learning.' There is essentially only one type . . . perceptual learning. In both cases this is the only sort of learning that needs to be assumed (6, p. 176)."

It seems difficult to reconcile with this statement the more rapid learning of Group W which has had *less* opportunity to structure perceptually the entire maze situation, i.e., fewer trials during the preliminary training stage.

In terms of reinforcement, the faster learning of Group W is quite meaningful. This group began the trial-and-error learning stage with weaker competing reaction tendencies. If we assume with Hull that the course of reinforcement is negatively accelerated, we would expect the earlier dominance of the correct response within Group W. The earlier increments of strength added by reinforcement to the correct reaction tendency within this group would be greater than those added to that reaction tendency within Group S. It follows that the correct response will occur with greater frequency.[3]

Summary

Sixty-four albino rats were employed in a simple T-maze under motivational conditions of hunger-food reward. Group W (32 animals) was subjected to a preliminary training period during which each animal was given four forced runs, two to the left and two to the right. Group S (a matched group) was given 40 reinforcements of the right-turning habit and 40 reinforcements of the left-turning habit. Both groups were then given a free-choice series of six trials to establish a baseline. Those animals that yielded a 6–0 or a 5–1 distribution of choices were discarded. Twenty-six animals of Group W and 21 of Group S met the criterion of response strength equality, a 3–3 or a 4–2 distribution of choices. The two selected groups were then introduced to the trial-and-error learning stage. Choice of one compartment was followed by reinforcement during the series and choice of the other was followed by a 10-sec. interval of confinement in the empty compartment (non-reward). During this stage, 48 trials were given, eight per night on six nightly sessions.

The learning curve (average percent of occurrence of the correct response as a function of number of trials) was seen to rise more rapidly for Group W. Statistically, a significantly greater total number of correct choices was made by Group W.

[3] This is rather facile reasoning. The reader is referred to the more rigorous and detailed derivation of Hull (5).

It was concluded that the speed of learning in a simple trial-and-error learning situation is some inverse function of the initial level of *absolute* strength that prevails among competing reaction tendencies.

(Manuscript received November 21, 1944)

References

1. Hull, C. L. Simple trial-and-error learning: a study in psychological theory. *Psychol. Rev.*, 1930, **37**, 241–256.
2. Hull, C. L. Mind, mechanism, and adaptive behavior. *Psychol. Rev.*, 1937, **44**, 1–32.
3. Hull, C. L. Simple trial-and-error learning: an empirical investigation. *J. comp. Psychol.*, 1939, **27**, 233–258.
4. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century Co., 1943.
5. Hull, C. L. *Elementary theory of individual behavior.* (Chapts. 3 & 4, manuscript on file, University of North Carolina Library.)
6. White, R. K. The case for the Tolman-Lewin interpretation of learning. *Psychol. Rev.*, 1943, **50**, 157–186.

AFFECTIVE DISTANCE AND OTHER FACTORS DETERMINING REACTION TIME IN JUDGMENTS OF COLOR PREFERENCE

BY WALTER C. SHIPLEY, JUDITH I. COFFIN, AND KATHRYN C. HADSELL

Wheaton College, Norton, Mass.

The object of this experiment was to apply improved psychological scaling methodology to an examination of the relationship between judgment time and affective distance in judgments of color preference. The specific problem was: *how does the time taken to choose the more pleasant of two colors vary with the affective distance between them?* This, of course, is a part of the broader problem of the role of *affective distance*, and the still broader one of the role of *psychological distance*, in the determination of judgment time.

Experimentation in this area has been handicapped by difficulties in scaling the purely psychological data with which it must deal. Somewhat recently Thurstone (**14, 15**), Guilford (**6, 7**), and others have developed a methodology for scaling such data in standard-deviation-type units. The present experiment was designed to apply this more meaningful and wieldy type of scaling to the above problem.

The literature concerning affective distance in relation to judgment time has been well reviewed by Dashiell (**3**) up to 1937, and will not be discussed here. The more recent literature will be considered in discussing our own findings.

The plan of the experiment was as follows. Reaction times were recorded for judgments of color preference made by the method of *paired comparisons*. A camera shutter simultaneously exposed the comparison stimuli and started an electric timer; reaction keys stopped the timer and indicated the choice. The choices were scaled for affectivity in standard-deviation-type units, and the affective distances involved in the respective choice combinations were plotted against the corresponding reaction times to reveal the functional relationship.

Subjects and Apparatus

The *Ss* were 40 women college students. Two additional ones who gave excessively slow sporadic reactions—reactions in excess of 10 sec.—were eliminated from the group.

The apparatus comprised a presentation and a registration unit. The presentation unit consisted of a 30 mm. camera shutter, adjustable to *S*'s height and opened and closed by *E*. Opening and closing was effected through the lever-action of a pushbutton-operated, electromagnetic stepping-relay to which the shutter-release was attached by a short cord. *S*, with her

head supported by a chin rest, viewed the color stimuli through the shutter aperture with her right eye. The stimuli were colored papers in the following six Milton Bradley hues: red, orange, yellow tint no. 1, green, blue shade no. 1, and violet shade no. 1. They were presented at eye-level behind two round holes, 2 cm. in diameter and 5.3 cm. from center to center, in a neutral gray cardboard screen 40 cm. behind the camera shutter. They were illuminated by partial daylight plus the light from a 60-watt bulb hidden from view, and a screen prevented their being viewed except through the shutter aperture; thus the stimuli appeared simply as two colored discs in a neutral gray field, with nothing else visible.

The registration unit consisted of a Springfield Timer and two reaction keys. The timer was started by the action of a voice-key clamped against, and activated by, the camera shutter; it was stopped by *S*'s pressing a reaction key—the key which indicated her choice.

The chief mechanical defects of the apparatus were its limitation of vision to the one eye, and a noisiness inherent in its shutter-opening mechanism.

Procedure

A. Experimental procedure

The stimuli were presented in accordance with the method of paired comparisons, in which each stimulus is presented in turn with each other one in pseudo-random order. Each *S* was given three such series in a single sitting—a practice series employing a different set of colors,[1] and two experimental ones. The second experimental series differed from the first in that the right-left relationship was reversed and the stimulus sequence was altered. As a further precaution against constant errors, four differing orders of presentation were used in the experimental series with 10 *S*s receiving each order. The stimulus combinations were presented about 10 to 15 sec. apart and were exposed until *S* had made a choice. A pause of about one min. was introduced between series.

The *S* was seated and asked to read the following instructions, after which the apparatus was adjusted to her.

> Sit in the chair and place your chin in the iron rest. Close your left eye. Look at the shutter with the right eye. Place the fore-finger of each hand lightly upon the key. When the shutter opens look at the colors and indicate the color you prefer by pressing the key on the side of that color. Be careful to keep your left eye closed tightly.

Nothing was said about speed or reaction time (although the operation of the timer was audible) and no ready signal was given.

Two *E*s operated the apparatus; one arranged the stimuli and recorded the choices by observing the reaction key, and the other set the timer and recorded the reaction times.

B. Scaling procedure

The stimuli were scaled on the basis of both color and choice (i.e., first choice, second choice, etc.).

Color scaling.—The scaling of the actual colors was not immediately relevant to the present problem, but the procedure is presented here to explicate the similar but more involved one of choice scaling; the obtained color scale values are offered in the hope that they may be of interest. The scaling was accomplished by finding the mean proportion of judgments favoring each color, converting these mean proportions to standard measurements, and adding a constant to eliminate negative values. The mean proportions were obtained by the following formula [2]

$$\text{Mean proportion} = \frac{C + N}{2nN}$$

with C representing the total number of preferences for any given color, n, the number of colors, and N, the number of judges. The resulting color scale is given in Fig. 1.

---

[1] Orange red, orange tint no. 1, yellow shade no. 1, yellow green, blue violet tint no. 1, and red shade no. 2.

[2] Modified from Guilford (7, p. 237) to fit the present double stimulus series.

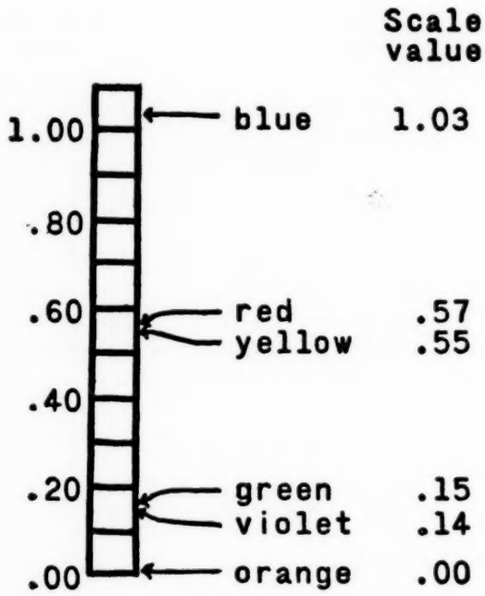

FIG. 1. Color scale.

Choice scaling.—The choices were scaled by the same general procedure, but with C in the formula representing the total number of preferences for a given *choice*, regardless of what the actual color was; for example, in arriving at the scale value for first choice, C represented the total number of preferences given the first choices of all Ss. The problem of tied-ranks, i.e., of equal preferences given two colors by the same S, was handled by eliminating the tied-rank data from the present computations. This reduced the Ns per choice by varying amounts from the possible total of 40. The resulting scale values, together with the Ns on which they are based, are presented in Fig. 2.

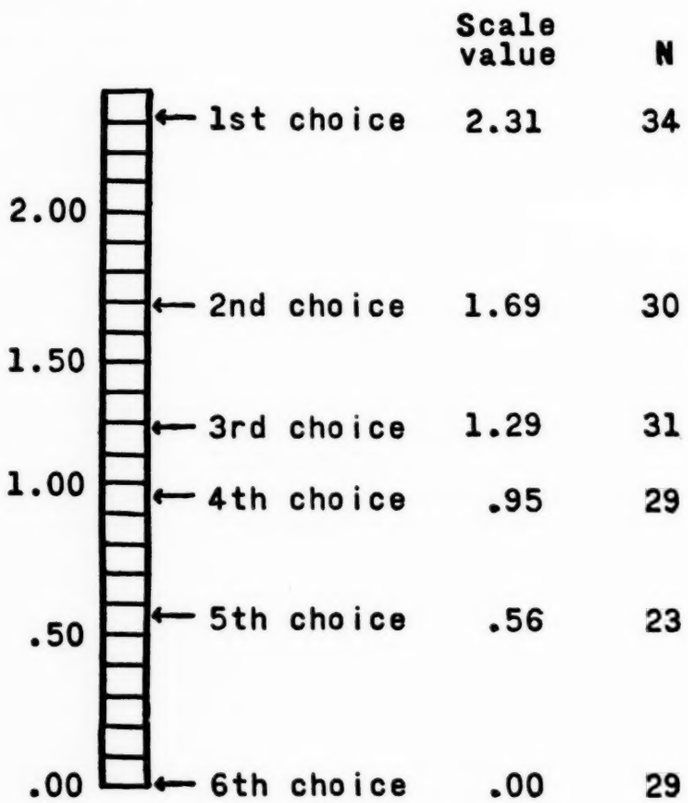

FIG. 2. Choice scale.

Results

A. Reaction time

The various Ss were found to differ considerably in mean reaction time. Their means, while averaging 1.41 sec., ranged from 0.74 to 2.61 sec., with a standard deviation of 0.44 sec., and single individual reactions ranged from 0.47 to 7.35 sec. In view of these differences it seemed desirable to reduce the data of all Ss to a common level, and with this purpose in mind each single reaction was converted to a percentage of the given S's mean reaction time. Only these converted scores were used for reaction time in the statistics to follow.

B. Reaction time in relation to affective distance

The general relationship between reaction time and affective distance is shown graphically in Fig. 3, in which mean reaction times

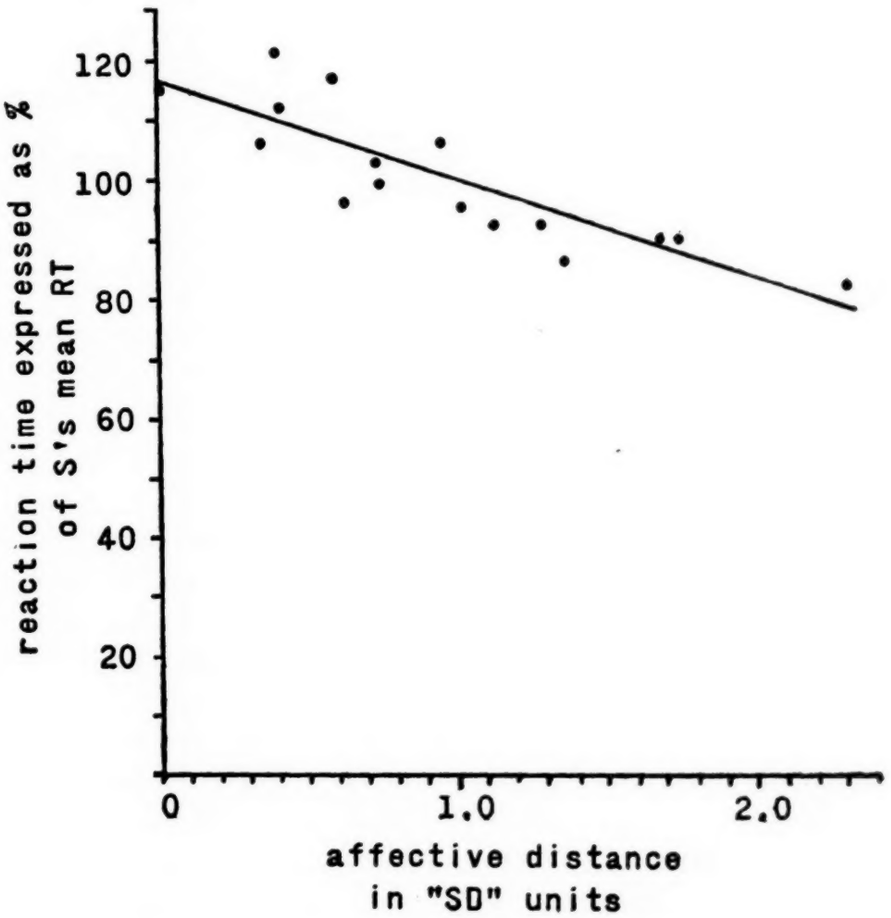

Fig. 3. Reaction time in relation to affective distance.

are plotted against affective distances. It will be seen that the quicker reactions are associated with the greater affective distances, the correlation for the plotted points being $r = -.86 \pm .045$, and

that the relationship approaches linearity. In view of this approximate linearity,[3] a straight regression line, computed by the method of least squares, has been drawn in.

The data from which Fig. 3 is plotted are given in Table I, which shows the mean reaction time, together with N and σ, for each choice combination. The tendency for quicker reactions to be associated

TABLE I

SHOWING MEAN AFFECTIVE DISTANCE, MEAN REACTION TIME, REACTION TIME σ, AND N FOR EACH CHOICE COMBINATION

Choice Combination	Mean Affective Distance	Mean Reaction Time *	Reaction Time σ †	N
1st & 6th	2.31	83	17.0	25
1st & 5th	1.75	91	18.1	20
2nd & 6th	1.69	91	21.6	23
1st & 4th	1.36	87	17.2	25
3rd & 6th	1.29	93	12.8	24
2nd & 5th	1.13	93	15.7	18
1st & 3rd	1.02	96	22.2	26
4th & 6th	0.95	107	28.5	20
2nd & 4th	0.74	100	23.6	21
3rd & 5th	0.73	104	28.8	18
1st & 2nd	0.62	97	25.5	29
5th & 6th	0.56	118	28.6	23
2nd & 3rd	0.40	113	24.3	26
4th & 5th	0.39	122	31.4	20
3rd & 4th	0.34	107	28.0	25
Tied ranks	0.00	116	24.0‡	25

* Computed from converted scores; each single reaction was converted to a percentage of the given S's average reaction time.

† The two scores per S were averaged before computing the σ's.

‡ Based on averages of from two to six scores per S.

with greater affective distances is highly reliable statistically, since the differences in mean reaction time for *any* of the five shortest, and *any* of the five longest affective distances (25 differences in all) are significant at the $P = <.05$ level, and all but three of them are significant at the $P = <.01$ level.

The coefficient of determination [4] of $d = .74$, arrived at by squaring the coefficient of correlation ($r = .86 \pm .045$) between the actual mean reaction times and those predicted on the basis of affective distance, indicates that about 74 percent of the obtained variance in mean reaction time was associated with variance in affective distance.

[3] The approximate linearity is evidenced by the comparability of η and r computed from the individual scores, η equalling $-.270 \pm .02$ and r, $-.264 \pm .02$.

[4] For a discussion of this measure see Guilford (7, p. 305 ff.).

C. Reaction time in relation to positive value of stimulus

It was observed that most of the points falling below the regression line in Fig. 3 represented stimulus combinations judged more pleasant, and those above it, combinations judged less so. This suggested the possible operation of a second factor, a positive-value-of-stimulus factor, over and above the affective distance one. To test this possibility a positive value score, the average of the choice scale values for the two stimuli, was assigned to each stimulus combination and then correlated with the algebraic deviations in mean reaction time from the regression line. The resulting correlation coefficient [5] was $r = -.70 \pm .09$, and the correlations between these deviations and the scale values for each of the two stimuli taken separately were $r = -.58 \pm .12$ for the more positive stimulus, and $r = -.59 \pm .12$ for the less positive one. Finally, the coefficient of determination for the positive value scores in relation to reaction time was $d = .16$. These facts taken together indicate the operation of this second factor,[6] reveal that it accounts for about 16 percent of the variation in mean reaction time, and suggest that the two stimuli contribute about equally to it. Its independence is indicated by the manner in which it was isolated and substantiated by a negligible correlation between positive value and affective distance of $r = .05 \pm .18$.

Discussion

In interpreting the results it should be borne in mind that the two factors revealed in the experiment were isolated in terms of mean scores. Their roles in determining single reactions, therefore, would be very much masked by chance variables. Bearing this in mind, we may proceed to a description of two closely related experiments, and then to a consideration of the factors themselves.

Both Dashiell and Barker, using the method of paired comparisons, have examined the relationship between choice-time and affective distance for judgments based on preference. Dashiell (3) used seven Milton Bradley colors as stimuli, presenting them in a Dodge tachistoscope and registering the choice times essentially as we have done. The colors were red, orange, yellow, green, blue, purple, and orange-yellow shade no. 2. Nine *S*s were used and the data were scaled by preference rank. In tabular form he presents mean times for each preference rank for each *S*, as well as for the group as a whole.

[5] The *N* for this and the following *r*'s was 15, the plotted point representing the tied-rank mean having been omitted.

[6] Confirmatory, though statistically nonsignificant, independent evidence of this factor was found in the tied-rank data. The mean positive value score for the 16 fastest of the 32 tied-rank reactions (the two reactions to a given combination having been combined for each *S*) was 1.18 as compared with 0.79 for the 16 slowest. The CR_σ for this difference was 1.95.

Barker (1) used seven liquids (beverages) as stimuli and recorded the choice times automatically on a moving tape by the tracing action of a lever which *S* moved in the direction of his preference. The *S*s were 19 boys, aged 9 to 11 years, and the liquids presented were orange juice, pineapple juice, tomato juice, water, lemon juice, vinegar, and saturated salt water. The choices were made under two conditions: a 'hypothetical' one in which *S* chose the drink which he thought he would prefer, and a 'real' one in which he was required actually to drink the liquids which he chose. Barker scaled his data in terms of preference rank and presents mean choice times, both for each *S* and for the group as a whole, for adjacent single-step ranks and for ranks separated by different step-distances.

Affective distance

Our finding that speed of judgment increases with the affective distance between the comparison stimuli is generally supported both by common sense expectancy and by the available experimental evidence in the affective field; it is similar to, and perhaps basically the same as, parallel findings for nonaffective psychophysical judgments. In the affective field it is firmly supported by the data of both Dashiell and Barker in the experiments just described, and by those of Godbeer (5), who found that when children were required to choose between a piece of candy and a quantity of toy soldiers regarded by them as being equal to it, they chose less rapidly than when this quantity was either increased or decreased. In the psychophysical field it is paralleled by the findings of Cattell (2), Henmon (8), and Kellogg (11), that speed of judgment increases with the objective distance between the comparison stimuli, and of Fernberger and Irwin (4, 10) that speed of judgment of singly presented (i.e., non-paired) stimuli increases as the stimuli deviate from the central tendency of the series.

As regards the actual *form* of the relationship between speed of judgment and stimulus distance, however, our results appear to be at variance with the other findings in that our data suggest a straight-line relationship, as contrasted with a curvilinear one for the others. Even when allowances are made for differences in scaling methods—allowances for conversion of physically scaled stimuli into psychological units (sense distances) and preference-ranks into standard deviation units—the discrepancy still seems to exist. The linearity of our curve may simply be the result of chance, or it may possibly be a function of some of the following factors—the narrowness of the range investigated, the specific part of the range investigated, or the amount of preliminary practice. The third possibility is supported by inspection of Barker's data (1) which reveal under both his 'real' and

'hypothetical' conditions a curvilinear relationship for the first series, but a much straighter one for the second. Unfortunately the data for our preliminary practice series, which would roughly correspond to his first series, are not available.

Positive value

Our finding that the positive value of the stimulus combination is also a determinant of judgment time is not so obviously predictable. It may even be spurious. Certain of the colors may have produced faster reactions *per se*, and through pure coincidence may also have been the most pleasing to the majority of *S*s. This possibility is rendered unlikely, however, by the fact that the mean reaction times for the colors, when plotted against the mean affective distances in which the colors were involved, showed only about half as much deviation from the regression line as did those for the choice combinations.[7] Thus, the lesser variability could hardly be invoked to account for the greater.

Furthermore, the finding is strongly supported by the previously described and very closely related experiment of Dashiell, and is partly supported by the somewhat less closely related, previously described experiment of Barker. Inspection of Dashiell's data reveals that his *S*s reacted more rapidly to the preferred choice combinations than to the unpreferred. In every case, nine in all, the mean reaction to the more positive combination was faster than to the symmetrically corresponding negative one; i.e., the mean reaction to combination 1–2 was faster than to 6–7, that to 2–5 was faster than to 3–6, and so on. This tendency, as computed by us, proved to be significant at the $P = < .01$ level. Barker found slower judgment times for negative alternatives than for either positive or neutral ones, but found little difference between these latter two.[8]

Finally, it is paralleled by still other experimental findings which perhaps may relate to it. In a choice discrimination experiment with rats, Schlosberg and Solomon (**12**) found that when animals trained to jump at a white card and ignore a black one were presented with choice discriminations on a white-black continuum, their speed of reaction was a function of the brightness of the more positive of the paired cards, and was relatively independent of the difference between them. In a simple reaction time experiment in which the stimulus

[7] The mean reaction times for the colors, when the tied-rank data were excluded, were: blue—92.9, yellow—97.0, orange—98.3, red—100.0, violet—103.0, and green—103.6, and the mean affective distances in which they were involved were, respectively, 1.16, 0.97, 0.99, 1.00, 0.98, and 0.93 units. The average of the mean reaction time deviations from the regression line was 2.53 units for the colors, as compared with 4.97 units for the choice combinations.

[8] Barker's stimulus series included some truly unpleasant stimuli, while ours probably did not.

was a sudden change in visual intensity, Steinman and Veniar (**13**) obtained consistently faster reactions to a constant stimulus ratio when the stronger of two intensities was employed; and, in a psychophysical experiment on comparison of linear magnitudes, Henmon (**8**) found that, while the mean reaction time to relative differences in linear magnitude was approximately constant, it decreased slightly as the total stimulus increased.

The actual nature of this positive-value factor is as yet undetermined. The question of whether it is an affectivity factor as such, or whether it relates simply to the basis on which the judgment is made, remains unanswered. As Hunt and Volkmann (**9**) point out, ". . . many phenomena in the field of affectivity can be explained by invoking the principles of judgment rather than the principles of a unique affective process." Thus this factor might well represent a tendency for *S* simply to react more rapidly as the stimuli possess that characteristic for which he has been instructed to look—in this case pleasantness. Repeating the experiment, but with *S* choosing the *less* pleasing member, should help to clarify this point. Should it prove, either in toto or in part, to be specifically affective in nature, further experimentation would still be required to determine its relationship to polarity of hedonic tone.

Summary

Judgment-times were determined chronoscopically for 40 *S*s as they indicated color preferences for six hues by the method of paired comparisons. Mean affective distances were computed in standard-deviation-type units and plotted against mean judgment times. While single reactions appeared to be governed largely by chance variables, mean speed of judgment was found to be determined chiefly by two factors:

(1) An *affective-distance factor* accounting for about 74 percent of the variability.

(2) A *positive-value-of-stimulus factor*, contributed to about equally by both comparison stimuli, accounting for an additional 16 percent of the variability.

(Manuscript received October 19, 1944)

References

1. Barker, R. G. An experimental study of the resolution of conflict by children: time elapsing and amount of vicarious trial-and-error behavior occurring. *In*: McNemar, Q., & Merrill, M. A. (Editors) *Studies in personality*. New York: McGraw-Hill, 1942.
2. Cattell, J. McK. The time of perception as a measure of differences in intensity. *Philos. Stud.*, 1902, **19**, 63–68.
3. Dashiell, J. F. Affective value-distances as a determinant of esthetic judgment-times. *Amer. J. Psychol.*, 1937, **50**, 57–67.

4. FERNBERGER, S. W., & IRWIN, F. W. Time relations for the different categories of judgment in the 'absolute method' in psychophysics. *Amer. J. Psychol.*, 1932, **44**, 505–525.
5. GODBEER, E. Factors introducing conflict in the choice behavior of children. Unpublished Master's thesis, Yale Univ., 1940.
6. GUILFORD, J. P. The method of paired comparisons as a psychometric method. *Psychol. Rev.*, 1928, **35**, 494–506.
7. GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1936.
8. HENMON, V. A. C. The time of perception as a measure of differences in sensation. *Arch. Phil. Psychol. Sci. Meth.*, 1906, No. 8.
9. HUNT, W. A., & VOLKMANN, J. The anchoring of an affective scale. *Amer. J. Psychol.*, 1937, **49**, 88–92.
10. IRWIN, F. W., & FERNBERGER, S. W. Einige zeitliche Bestimmungen der Urteile in der Methode der einzelnen Reize. *Arch. ges. Psychol.*, 1933, **89**, 541–548.
11. KELLOGG, W. N. The time of judgment in psychometric measures. *Amer. J. Psychol.*, 1931, **43**, 65–86.
12. SCHLOSBERG, H., & SOLOMON, R. L. Latency of response in a choice discrimination. *J. exp. Psychol.*, 1943, **33**, 22–39.
13. STEINMAN, A., & VENIAR, S. Simple reaction time to change as a substitute for the disjunctive reaction. *J. exp. Psychol.*, 1944, **34**, 152–158.
14. THURSTONE, L. L. A law of comparative judgment. *Psychol. Rev.*, 1927, **34**, 273–286.
15. THURSTONE, L. L. The method of paired comparisons for social values. *J. abn. soc. Psychol.*, 1927, **21**, 384–400.

AN EXPERIMENTAL ANALYSIS OF DYNAMIC AND STATIC EQUILIBRIUM

BY ROLAND C. TRAVIS

Stanford University

Historically, postural balance has been divided into two related although different processes, namely, static equilibrium (maintenance of relatively fixed postures) and dynamic equilibrium (act of establishing equilibrium or gaining bodily orientation in relation to gravity). Static equilibrium requires a continuous tonic contraction of certain skeletal muscles, e.g., in sitting and standing. Dynamic equilibrium is the reorientation of the body or reëstablishment of equilibrium after the body has been thrown off balance in relation to gravity. Also, the bodily adjustments and behavior associated with simple rectilinear and rotary accelerations often resulting in vertigo, dizziness, and nystagmus are phenomena of dynamic equilibrium. Furthermore, certain behavior patterns in locomotion (running, jumping and the like) where bodily orientation is maintained, despite interrupted contacts with the ground or fixed surfaces, may be considered important aspects of dynamic equilibrium.

Various sensory modalities, vision, audition, kinesthesis, vestibulation, the skin senses and the visceral senses give in different ways essential information in all postural adjustments. It is impossible adequately to attribute dominance to any one sense field over the others in posturing reactions in view of the necessity for the combination of sensory data from a number of sense fields in this type of sensori-motor adjustment of the body as a whole.

The distinction between static equilibrium and dynamic equilibrium in man presumably rests upon the differences between reflex, ballistic motor patterns on the lower brain and spinal cord levels and the fine coördinated motor patterns on the mid-brain and cortical levels. This distinction is the logical result of the viewpoint of 'hierarchial functioning' of neural centers from the simplest to the most complex levels. From the viewpoint of 'unitary functioning' of the nervous system the distinction between static and dynamic equilibrium is more apparent than real.

The present study is an experimental analysis of various behavioral aspects of postural balance and body orientation, including an analysis of individual and sex differences with respect to these traits. Some of the more important problems may be stated by specific questions.

(1) What effect does practice have on balancing skill? (2) What is the relative importance of the visual sense fields in postural balance? (3) Does exercise effect postural balancing skill? (4) What is the correlation between static equilibrium as measured by the ataxiameter and dynamic equilibrium as measured by the stabilometer? (5) What are the relationships between the physical measurements of weight, height, and distance from the center of gravity of the body to the standing base, length of feet and balancing skill? (6) What are the relationships between postural balance as measured by the dynamic stabilometer and orientation of the body as measured by rotation experiments? (7) What are the relationships between perceptive abilities and motor abilities in postural balance and body orientation? (8) What is the extent of the variability within the individual and between individuals in postural balance and body orientation? (9) What are the relationships between stabilometer performance, manual pursuit, steadiness and reaction time?

These questions cover a variety of behavior patterns resulting from a number of different stimulating situations, yet fundamentally they are quite closely related and interdependent.

Effect of Practice, Mild Exercise and Visual Cues on Postural Balance

A. Stabilometer results.

In a former study (1) by the author on dynamic equilibrium a new stabilometer was described in detail, and the reader is referred to that report for the details of construction, standardization and analysis of procedure. This new stabilometer presented a real challenge to the individual in demanding voluntary efforts in a finely coördinated fashion in the maintenance of postural balance while standing on an unstable platform. The *S* was instructed to stand on the platform, feet together, hands at the sides, and attempt to keep the stylus on the target. The stylus which was fastened to an offset from the platform moved over a small brass target. The stylus and target were in circuit with an interrupter and electric counter which scored continuously if *S* was successful in keeping the stylus on the target. Any unsteadiness in excess of two degrees in attempting to keep the platform balanced broke contact between the stylus and target thus interrupting the score. A perfect score was 300 in one min. Each *S* was permitted a preliminary trial to adjust to what he considered the most favorable position on the platform with the restriction that the toes and heels be kept together. The experiments were conducted with shoes except where specific mention is made to the contrary.

To determine the influence of visual cues on balancing skill a study of 22 college men and 22 college women was made. Each *S* was given eight one-min. trials in the situation with visual cues with rest periods of one min. between trials and after a few minutes' rest eight one-min. trials blindfolded with rest periods of one min.

Fig. 1 shows graphically the means and standard deviations for the men and the women separately for the two situations with visual cues and without visual cues for eight trials in each situation. It will be noted: that there is a small percentage gain (about seven percent) from the first two trials to the last two trials for the situation with visual cues and four percent without visual cues for men and women

combined; that for both men and women there is a large difference between the average scores for the two situations; and that the women stood much higher than the men.

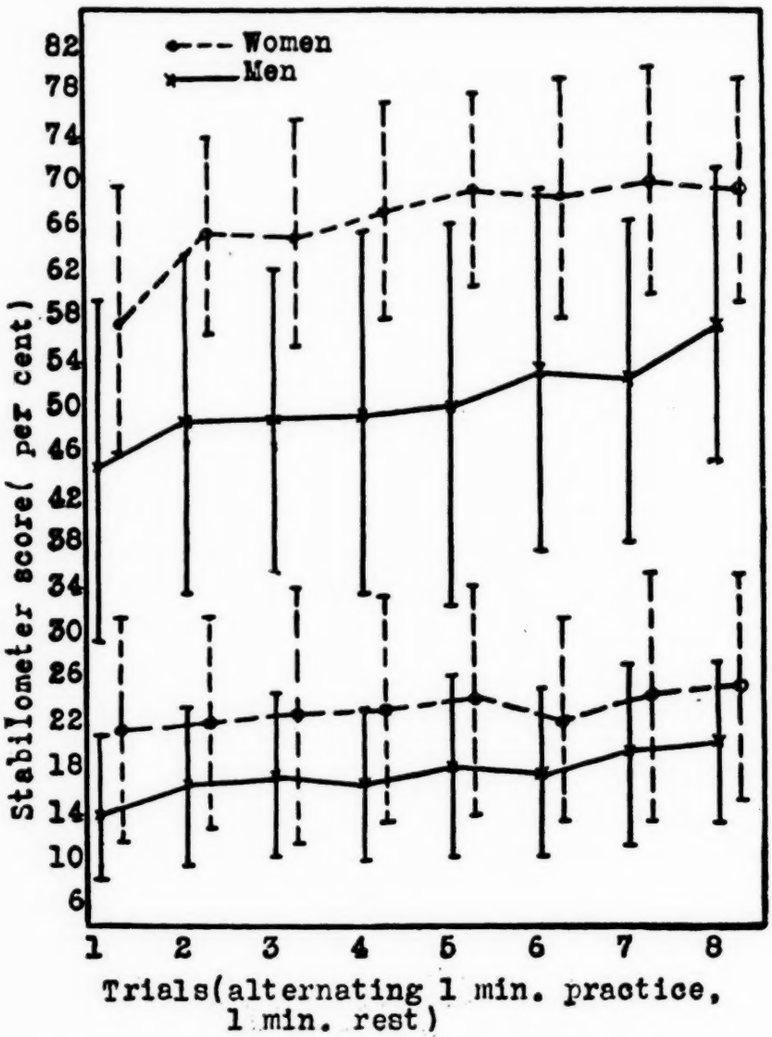

FIG. 1. Stabilometer scores for successive trials with and without vision. Means (horizontal heavy and broken lines) and standard deviations (vertical heavy and broken lines) for men and women separately. Two top curves represent performance with vision, two bottom curves without vision. (N = 22 men and 22 women.)

Despite the fact that there is slight improvement on the average as trials continue, a few cases almost doubled their scores from the first to the last trial, and a few others showed a decrease in successive trials. On the whole the learning factor seems to be relatively unimportant during this series of trials for most *S*s.

To determine the effects of more extensive practice and mild exercise on the stabilometer scores another study of 17 college students was made. The results of this study are presented graphically in Fig. 2. Trial-series 1 and 2 were given the first day, two days later trial-series 3 and 4, and two days later trial-series 5 and 6. Trial-series 1, 3, and 5 constituted three one-min. trials each with rest periods of one min.

On the first day trial-series 1 was followed by trial-series 2. Trial-series 2 consisted of three one-min. trials, each of which was preceded by an exercise period. Each *S* was required to jump up and down 10 times on each foot in succession just before each individual trial in trial-series 2, and no rest periods were permitted. The same procedure was followed for all three days.

The main point to be made here is that the means for this entire series indicate a small gain equivalent to that shown in Fig. 1. The effect of exercise on stabilometer scores is variable between individuals. For some *S*s this amount of exercise constitutes a 'warming-up' process with consequent improvement in score, while others

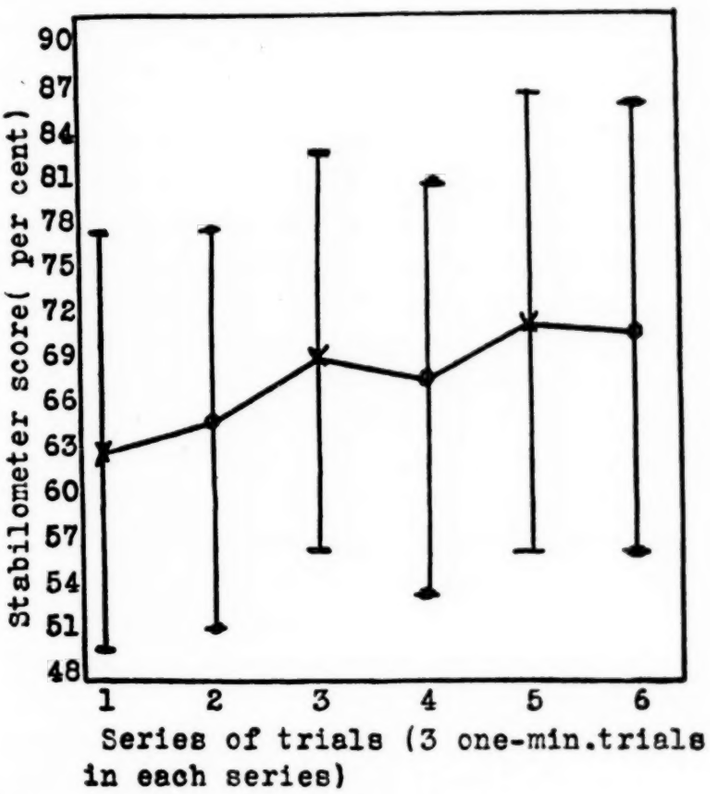

FIG. 2. Stabilometer scores for three different days before and after exercise. Series 1, 3 and 5 were given before exercise; series 2, 4 and 6 after exercise. Horizontal line indicates mean scores for each series. Standard deviations are shown by the vertical lines for each series. ($N = 17$.)

become somewhat fatigued with dropping-off in scores. For the group the exercise trials show an improvement over the normal trials for the first day but a decrement for the second and third days. However, on the whole the general trend is up.

Table I summarizes the data from another study of forty-two college men and forty-two college women on the stabilometer alternating three one-min. trials with visual cues with three one-min. trials blindfolded with one-min. rest periods between trials. The trials with and without vision in this study were alternated to control practice and fatigue effects.

The importance of vision in postural balancing skill is apparent. The performance in the situation with visual cues is almost three times as effective as with the blindfold technique. The sensory data avail-

TABLE I

STABILOMETER SCORES (PERCENT). MEN AND WOMEN COMBINED

	With Visual Cues	Blindfolded
N	84	84
Mean	56	20
Sigma	16.4	11.8
Sigma (mean)	1.8	1.3
Sigma (diff.)	1.5 (for correlated data, $r = .60$)	
Critical ratio	24	

able to the individual common to both the seeing and blindfold situations are kinesthetic, tactual, vestibular, organic, and auditory cues. The essential difference is the presence of vision in one situation and not in the other. It is instructive to add that the extension springs tend to keep the platform in a level position but the live weight of a standing person imposed upon it with a continuously shifting center of gravity (**2**, **3**) produces a more or less continuous movement of the stylus across the contact-target. As long as the stylus remains on the target the clicking of the counter is clearly audible to the *S*. Therefore, in addition to the somatic sensory data as to whether the platform is on-balance or is off-balance, the audible clicking gives the *S* further information on the effectiveness of his performance. In the situation with vision, not only is *S* able to make use of fine visual cues (observing the stylus and target) but in addition vision seems to facilitate the whole performance by a more effective integration of the other sensory data, which is not possible when visual cues are absent. The results of these two studies as presented in Fig. 1 and Table I correspond closely in pointing out the great importance of vision in postural balance of the dynamic type.

Most of the differences between the sexes as shown in Fig. 1 is explained on the fact that weight and height correlate negatively with the stabilometer score and the men on the average are taller and heavier than the women. These factors are analyzed further in a later section.

B. Ataxiameter results.

One aspect of static equilibrium in man, namely, sway, has been measured by the ataxiameter while the individual was in the standing position (Miles, **4**; Fearing, **5**; and Edwards, **6**). In these studies as well as in the present study sway was measured in mm. in the lateral and anteroposterior planes under a variety of conditions with the eyes open and with the eyes closed.

The Miles ataxiameter (**4**) with certain modifications was used in the present study. A short ribbon with a small weight at each end was suspended over the axle of each wheel-adder and served to indicate whether the wheel had rotated more than one revolution. The pawls came in contact with a thin rubber band placed over the outer surface of each wheel permitting the wheel

to turn in one direction only. The cone bearings of each wheel were adjusted carefully to reduce friction to a minimum. Each cord (four in all), passing from the headpiece and over the grooved pulley of each wheel-adder, was kept taut by suspending a small weight from the farther end. To secure reliable results with this apparatus it is essential to have positively acting pawls to permit rotation of the wheel-adder in one direction, to reduce friction of the wheels to a minimum, and to have a reliable check on whether the wheels have rotated more than one revolution. Frequent checks on the extent of movement of the cords and the corresponding extent of movement of the wheels showed an average error of about one percent. With careful checking and proper adjustments the reliability of this technique as measured by correlating successive trials was increased from about .47 to .85.

Each *S* was instructed to stand firmly on the floor, hands at the sides, feet together, and to align the small white bead (18 in. in front of the eyes) on a white circle eight feet beyond, using the dominant eye. The subordinate eye was covered. *E* adjusted the *S* with the headpiece in place so that he stood in the center of the square frame which held a wheel-adder at each corner. The white bead was adjustable for different heights. After the *S* was adjusted and standing comfortably he was instructed to stand as quietly as possible for one min. and the trial was begun.

Three one-min. trials were given with the dominant eye aligning the bead on the circle and three one-min. trials with the eyes closed with rest periods of one min. between all trials. During the rest period the *S* was seated. The trials for the seeing situation and the blindfold situation were alternated to control learning and fatigue effects. These two situations were followed by three more one-min. trials, each of which was given right after a short exercise period. The exercise constituted hopping up and down 10 times on each foot in succession.

TABLE II

DISTRIBUTIONS OF SWAY SCORES FOR THE THREE SITUATIONS: DOMINANT EYE OPEN; EYES CLOSED; TRIALS AFTER EXERCISE

Sway in mm.	Dominant Eye Open Frequency	Eyes Closed Frequency	After Exercise Frequency
1040		1	
1000			
960			
920			
880		1	
840			
800			
760			
720		1	
680		2	
640		1	
600	1	2	1
560		8	1
520		2	
480		5	
440	2	7	2
400		6	1
360		3	1
320	3	7	7
280	12	2	10
240	10		11
200	11	1	9
160	6	1	3
120	5		4
N	50	50	50
Mean	260	490	285
S.D.	87	152	96
$CR_{1-2} = 16$		$CR_{1-3} = 4$	

Table II gives the distribution of average sway scores in mm. for three trials for the three situations. Four college men and 46 college women took part in this study. For the dominant eye open the mean sway was 260 mm. with an S.D. of 87 mm.; for eyes closed the mean was 490 mm. and an S.D. of 152 mm.; for the trials after exercise the mean was 285 mm. with an S.D. of 96. The ratio of the means to their standard error for the dominant eye open and the eyes closed was 16; it was 4 for the means of the dominant eye open and trials after exercise.

With the dominant eye open the sway score was 47 percent less than with the eyes closed and the variability was a little more than one-half as great.

Table III gives the trial to trial reliability *r*'s and the intercorrelations of the sway scores in the three situations. As will be noted the

TABLE III

TRIAL TO TRIAL COEFFICIENTS AND INTERCORRELATIONS OF THE SWAY SCORES FOR THE THREE SITUATIONS. $N = 50$

	Trial 1 with Trial 2	Trial 2 with Trial 3	Trial 1 with Trial 3	Eyes Closed	After Exercise
Dominant eye open	.74	.87	.77	.80	.90
Eyes closed	.85	.70	.48		
After exercise	.92	.92	.83	.66	

reliability *r*'s are .70 or above, except for Trial 1 with Trial 3 for the eyes closed. The 'after exercise' trials are most consistent and seem to indicate a stabilizing effect of mild exercise on steadiness in standing. The relatively high *r* between the dominant-eye-open situation and the trials-after-exercise situation would be expected in that the only apparent variable was the exercise.

There is a small increase in sway after the first trial for the *dominant-eye-open* situation, a small decrease for the *eyes-closed* situation, and an increase for the *after-exercise* situation. Some variability from trial to trial is evident, and the only consistent rise is during the 'after-exercise' trials (Table IV).

TABLE IV

MEAN SWAY SCORES OF SUCCESSIVE TRIALS IN THE THREE SITUATIONS

	Trials			Average
	1	2	3	
Dominant eye open	253	264	262	260
Eyes closed	499	479	493	490
After exercise	274	284	298	285

Fearing (5) reported an average reduction in the sway scores with the eyes closed of about 14 percent for 10 *Ss* from the first two trials to the last two trials in a series of 10 two-min. trials on different days. Some learning seems to result in the 'eyes-closed' situation whether the rest periods are long or short.

Preliminary observation indicated that breathing seemed to increase the amount of sway as measured at the head. During inspiration the chest moves outward, and the head moves backward to compensate for this disturbance in equilibrium. During expiration the reverse relationship holds. To test this hypothesis the procedure of alternating normal half-min. trials with half-min. trials during which the *S* held his breath was adopted with rest periods of one min. between trials. Half-min. trials were chosen because of difficulties in voluntarily holding the breath longer than this. Three normal-breathing trials were alternated with three holding-breath trials for 26 *Ss*.

The average sway for the normal breathing trials was 333 mm. and for the holding-breath trials, 320 mm., a difference of 13 mm. or 4 percent, which gives a standard critical ratio of 1.24. A tendency toward decreased sway in the holding-breath trials is apparent but is not very significant. However, as exercise continues, respiration increases in rate and magnitude, thus increasing the head movement, and in turn increasing the amount of sway.

It is apparent from the data presented in Table II and Table IV that a fine visual point of reference using the dominant eye is of great importance in controlling sway as compared to the situation with the eyes closed. Furthermore, sway is increased significantly (about 10 percent) after short periods of exercise. The increase in sway after exercise is partially if not wholly explained on the basis that respiration causes a slight head movement in the anteroposterior plane, and as exercise causes respiration to increase in frequency and magnitude, an increase in sway would thus be expected.

Intercorrelations Among Weight, Height, Foot Length, Stabilometer Scores, and Sway Scores. Summary of Four Different Studies

Weight was measured by a standard laboratory balance scale in pounds, height by a Stoelting stadiometer in inches. Length of the feet was measured by drawing an outline of each foot on a sheet of paper placed on the stabilometer and taking the longest distance of the longer foot in mm. The stabilometer scores and the sway scores are the average of three one-min. trials. The sway scores were obtained a few min. after the stabilometer scores.

Table V gives the correlations between stabilometer scores, sway scores, the factors of weight, height, and foot length and trial to trial scores.

TABLE V

Intercorrelations

	Study 1 $N = 67$	Study 2 $N = 50$	Study 3 $N = 50$	Study 4 $N = 40$
1. Sway with height	−.06	.15		
2. Sway with weight	−.09	.06		
3. Sway with foot length	−.05			
4. Weight with height	.78		.70	.66
5. Weight with foot length	.70			
6. Height with foot length	.80			
7. Stabilometer exercise with wgt.			−.67	
8. Stabilometer with weight	−.75		−.83	
9. Stabilometer exercise with hgt.			−.61	
10. Stabilometer with height	−.58		−.61	
11. Stabilometer with foot length	−.48			
12. Stabilometer with sway	−.04			−.10
13. Stab. exer. with sway exer.				−.03
14. Stab. normal with stab. exer.			.92	
Reliability Coefficients				
1. Sway trial 1 with trial 2	.67	.74		
2. Sway trial 2 with trial 3	.54	.87		
3. Stab. exer. trial 1 vs. trial 2			.90	
4. Stab. trial 1 with trial 2	.82		.91	
5. Stab. exer. trial 2 vs. trial 3			.91	
6. Stab. trial 2 with trial 3	.91		.89	
7. Stab. trial 1 with trial 3			.87	
8. Stab. exer. trial 1 with trial 3			.89	
9. Eyes open vs. eyes closed (sway)		.80		
10. Eyes open vs. eyes closed (after exer., sway)		.90		.90

From these correlations the statement that height, weight, and foot length have no effect on static equilibrium but seem to be of considerable importance in dynamic equilibrium seems warranted. Furthermore, static and dynamic equilibrium are unrelated (item 12). The reliabilities for the ataxiameter are not as high as those for the stabilometer. The correlations between weight and height and foot length are typical of those previously found. Harris (7) reported two correlations of .73 and .75 of two independent studies between height and weight, and .71 and .73 between height and foot length.

In view of the fairly high negative correlations between stabilometer scores and the factors of weight and height and a fairly high positive correlation between weight and height, the question arises as to whether weight or height is the more important factor in affecting the stabilometer score.

The partial r between weight and the stabilometer score with height constant resulted in a negative coefficient of −.61, and that between height and stabilometer score with weight constant resulted in a negative coefficient of −.10 for 164 men. The same partial r's

for 157 women were −.70 and +.16. From these partial correlations it would seem that increased weight is the more important factor in producing a low score on the stabilometer.

Another question arises, as to the effect of the distribution of weight throughout the body axis. For example, would the person with a relatively larger proportion of weight in the chest, shoulders, and head regions be more unstable on the stabilometer than one with a relatively larger proportion of weight in the lower extremities? Several suggestions present themselves in answer to this question. First, we may determine the *r* between the height-weight ratio and stabilometer score. Second, we may select those cases identical in height and correlate weight with stabilometer score. Third, we may select those cases identical in weight and correlate height with stabilometer score. Fourth, we may correlate the distance from the center of gravity of the longitudinal axis of the body to the standing base with the stabilometer performance.

Let us consider these suggestions in turn. The Pearson *r*'s between height/weight ratio and the stabilometer score were .64 for 167 men and .70 for 173 women. These *r*'s seem to indicate a definite tendency for the stabilometer scores to increase as weight decreases for a given height, or, a definite tendency for the stabilometer scores to increase as height decreases for a given weight. The first conclusion is borne out by the partial *r*'s, but the second hypothesis contradicts the previous result with the partial correlation technique.

A more direct answer to this question is presented in Tables VI, VII, and VIII. Table VI gives the correlation between weight and

TABLE VI

CORRELATION BETWEEN STABILOMETER SCORE AND WEIGHT AT VARIOUS HEIGHTS

Heights Inches	*r*	*PE*	*N*	Weight Range
Men				
67 & 68	−.62	.08	31	124–172
68.5–69.5	−.73	.06	26	130–186
70	−.67	.09	22	133–185
71	−.69	.07	26	125–195
72	−.24	.12	30	141–189
73 & 74	−.50	.10	23	136–188
Average *r* = −.58				
Women				
64	−.82	.04	27	104–160
65	−.92	.02	37	100–170
66	−.46	.08	45	108–160
67	−.68	.06	39	103–167
68	−.78	.05	26	115–175
Average *r* = −.73				

TABLE VII

CORRELATION BETWEEN STABILOMETER SCORE AND HEIGHT WITH RESTRICTED WEIGHT GROUPS

Weight Groups	*r*	*PE*	*N*	Height Range
Men				
147–153	−.10	.12	28	65.5–74.5 in.
157–163	+.34	.12	25	65–74 in.
Average *r* = +.12				
Women				
122–126	−.08	.20	10	63.5–68.2 in.
127–132	−.07	.11	35	63–68.5 in.
138–142	−.26	.13	26	62–71 in.
Average *r* = −.16				

TABLE VIII

EFFECT OF HEIGHT WITHIN RESTRICTED WEIGHT GROUPS ON STABILOMETER SCORE

Weight Groups	*N*	Height Range (In.)	Score	Difference
Men				
147–153	19	65.5–69.5	54	1
	9	70–74.5	53	(favor of short)
157–163	10	65–69	47	2
	15	70–74	49	(favor of tall)
Women				
127–132	9	63–65.5	70	4
	26	66–68.5	66	(favor of short)
138–142	11	62–66	60	1
	15	67–71	61	(favor of tall)

stabilometer scores for various samples of men and women at definite heights. The average *r* is −.58 for the men and −.75 for the women. These *r*'s corroborate those in Table V in indicating a definite tendency for an increased stabilometer score with a decrease in weight

The low *r*'s in Table VII between height and stabilometer score with weight restricted corroborate the partial *r*'s in indicating that height as such seems to have little effect on the stabilometer score. The results in Table VIII also confirm this conclusion by the small insignificant differences between the tall and the short in each weight group for both men and women.

The distance from the center of gravity (CG) of the longitudinal axis of the body to the base line of the feet was determined in the following manner. A wooden platform seven feet in length and two feet in width was mounted horizontally on knife edge bearings to support a reclining person in a manner to produce a balanced teeter-totter. A small sliding weight with an adjusting screw made possible the maintenance of a perfect balance of the platform when not in use. Pre-

liminary observations indicated that the CG-distance was greater than one-half the height of an individual; thus the knife edge bearings were mounted slightly off-center with a counter-balancing weight attached to the shorter or head end. The knife edge bearings were fastened to the platform in a manner to produce an underslung status of the platform. This permitted the axis of rotation of the platform to coincide (to a close approximation) with a line through the center of the body in the lateral plane.

While reclining on the platform *S* can shift his position and weight along the long axis of the platform until he reaches a balanced position. The shifting of *S*'s position was aided by a longitudinal board one by four inches fastened on each side of the platform to which *S* could grasp with his hands, and so move his whole body through small distances until he was perfectly balanced. *S*'s CG-distance was measured after he was balanced by reading the distance on the scale (fastened to side of platform) in in. to the nearest tenth from the center of gravity of the platform to the base line of his feet. The base line of his feet was determined by a double T-square. When properly aligned the T-square assured a vertical position of the feet. Care was taken to have *S*'s hands at his sides and his feet in a vertical position when a balanced status was accomplished.

Even though *S* was required to lie quietly, hands at the sides, feet in a vertical position, some adjustment was always necessary after a seemingly balanced condition. For example, the movement of the toes one-fourth in. toward or away from the head would result in an imbalanced status of *S* and the platform, indicating a high degree of sensitivity of this method. With a little practice, however, *S* was able to establish a balanced position within a few seconds. Repeated readings were usually identical with an occasional variation of one to two-tenths of an in. The platform and bearings were mounted on a long table 28 in. high with blocks between the ends of the platform and the table to restrict the teeter-totter movement.

The factors measured in this particular study (in stocking feet) were the CG-distance, weight, standing height, sitting height and balancing skill on the stabilometer. Seventy university students (27 men and 43 women) acted as *S*s. Table IX gives the various data on the relevant components in correlating CG-distance with stabilometer performance.

TABLE IX

THE MEANS, SIGMAS, RANGES AND INTERCORRELATIONS OF RELEVANT COMPONENTS IN RELATING CG-DISTANCE WITH STABILOMETER PERFORMANCE

	Mean	S.D.	Range	Correlations							
				1	2	3	4	5	6	7	8
1. Weight (lbs.)	140	23.2	96–211		.72		.66	−.67			.00
2. Height (in.) standing	66.7	3.5	61–76			.85	.89				
3. Height (in.) sitting	34.7	1.4	32–39		.85		.77				
4. CG-distance (in.)	38.2	2.5	34–45		.89	.77		−.61			.00
5. Stab. score (percent)	58	17	22–94	−.67			−.61		.76	.12	.00
6. CG/wgt. × 100	28	3.6	19–37					.76			
7. CG/hgt. × 100	58.7	1.15	54–62					.12			
8. Age	18.2	1.7	17–24	.00			.00	.00			

It is evident from Table IX that height, weight and CG-distance represent a complex of body characteristics which seems to be fairly consistent for these 70 men and women. The *r* of .89 between CG-distance and standing height indicates that the distribution of weight throughout the longitudinal axis of the body is quite uniform for this group. Perhaps the restricted age range (17 to 24 years) would be important in this connection.

Apparently CG-distance as a measure adds little to the analysis of the effect of physical characteristics on balancing skill for this group.

The ratio of CG-distance/wgt. has a fairly high correlation, whereas the ratio of CG-distance/hgt. has a low r with the stabilometer score, indicating again the importance of weight and the relative unimportance of height in balancing skill of this type.

In this study, age (within the range 17 to 24 years) was unrelated to weight, CG-distance and stabilometer score. However, in Study 3, Table V, the r's between age (15 to 26 years) and stabilometer score, age and weight, and stabilometer score and weight were $-.56$, $.55$ and $-.85$ respectively. The partial r for age and score with weight constant was $-.24$, and for score and weight with age constant r was $-.80$. Here again weight was the important factor.

Table X gives the relevant data in the determination of the sex difference in balancing scores on the stabilometer. The weight groups

TABLE X

A. Stabilometer—Sex Difference with Weight Controlled

	N	Mean	S.D.	S.D.M.	S.D. Diff.	C.R.
Women	61	59	11	1.4		
Men	43	55	9	1.4	1.95	2.0

B. Weight groupings for determining sex differences.

	Weight Group	*N*	Weight Mean (lbs.)	Height Mean (in.)	Score
Men	140–144	9	141.4	70	57
	145–149	15	147	68.4	54
	150–154	19	151	69.5	55
	Average Score = 55				
Women	140–144	26	142	68	61
	145–149	18	147	66.6	61
	150–154	17	152	67.6	53
	Average Score = 59				

are definitely restricted to control the weight difference between men and women. In the same weight group, however, the men on the average are taller by about two in. than the women. The average difference in stabilometer score is four percent with a C.R. of 2.0. A definite trend is apparent here, but it is not very significant statistically.

In view of the facts (1) that weight seems to be of great importance in postural-balance skill of this type whereas height has little influence, and (2) that there is a small insignificant sex difference, it seemed advisable to include distributions of stabilometer scores for all the different weight groups for men and women combined. Table XI

TABLE XI

DISTRIBUTION OF STABILOMETER SCORES (PERCENT) FOR DIFFERENT WEIGHT GROUPS, MEN AND WOMEN COMBINED

Actual means, means corrected for curve of 'best-fit' and sigmas for each weight group are given.

Stabilometer Score (percent)

Weight (lbs.)	10	14	18	22	26	30	34	38	42	46	50	54	58	62	66	70	74	78	82	86	90	94	*f*	*M*	M_e	S.D.
215		1																					1			
210																										
205						1																	1			
200																										
195						1	1																2			
190	1	1		1			1	1	1														6	27	28	12
185	1		2	1	1	2	3	1	1	1	1												14	31	31	11
180							2	2	2	1	2												9	42	34	6
175					3	1	3	1	3	2	1	1	2										17	40	38	10
170		3				3	1		2	1	1	1											12	33	41	13
165			1	1	2	2	1	4	4	2	4	4	2	2									29	43	44	11
160					2	1	1	2	5	7	8	4	1		1								32	46	47	8
155					1	1	2	3	7	4	4	4	3	3	1	2	1						36	49	51	11
150							3	2	4	1	4	7	4	8	2		1						36	53	54	10
145						1		1	1		4	9	8	5	1	1	1	1					33	56	57	9
140							1		2	4	2	3	7	6	3	1	5						34	58	61	10
135									1		3	3	4	4	9	4	3	1					32	63	64	9
130										2	1	4	9	7	8	6	7	4					48	64	67	8
125											1	4		6	1	8	5	7	2				34	69	70	9
120													1		1	8	4	9	5				28	75	73	6
115												1			2	4	4	8	4	1			24	75	76	7
110															2	2	1	2	4	5	1		17	80	79	7
105																		4	4	4	3		15	83	82	4
100																1			1	2	2		6	84	86	7
95																			1				1			
90																						1	1			

	Weight (lbs.)	Stabilometer Score (percent)
Mean	144.6	60.2
S.D.	22	16.8

Total N = 468 (181 men, 287 women)
Pearson r between weight and score = −.83

represents these distributions with the actual means and S.D.'s for each weight group and the means as determined from the curve of 'best fit,' assuming a linear regression of weight on stabilometer score.

It is obvious from Table XI that a relatively high negative correlation exists between the stabilometer performance and weight, that great variability between individuals for each weight group occurred, and that considerable variability in S.D.'s from weight group to weight group also occurred.

CORRELATION OF PERFORMANCE ON THE STABILOMETER WITH PERFORMANCE ON THE BODY ORIENTATION APPARATUS

In view of the necessity for interplay of fine voluntary postural adjustments and tonic reflex mechanisms in the performance on the stabilometer, the question arose as to whether the individual's ability to coördinate voluntary and reflex adjustments of this character would be related to the task of a continual voluntary reorientation of the body such as is required in the performance on the double-axle rotation chair. The latter task is predominantly voluntary in character,

although there are certain postural reflexes which are sometimes in harmony with, and sometimes antagonistic to, the essential voluntary responses.

The technique, procedure, and standardization of the double-axle rotation chair were described in a previous report (**8**) and only the main essentials will be given here.

The rotation chair is mounted on two automobile wheel hubs, the axles in line, one above the other. The chair is fastened to the top hub, the driving mechanism to the lower hub. The chair and its hub rotate on the lower hub by moving a lever which connects the two hubs. Extension springs also connect the two hubs so that as the lower hub is rotated, the upper hub and chair also are rotated. By appropriate movements of the lever the *S* is able to keep the chair stationary in space while the lower hub is being rotated clockwise and counterclockwise by the driving mechanism.

The driving mechanism consists of a multiple-lobe cam which is rotated either clockwise or counterclockwise by an electric motor through a speed reducer. The follower-wheel which impinges on the cam is fastened to the driving rod, which in turn is fastened to the lower hub of the rotation chair. The eight cam lobes produce a variety of acceleration rates of movement of the chair. The adequacy of performance is recorded by an electric interrupter and counter system.

For studying the perceptual aspects of body-rotation simple electric response keys and motor counters were used instead of a control lever.

The reliability coefficients for the body orientation tasks, both for the perceptual and motor components, are given in Table XII. The

TABLE XII

RELIABILITY COEFFICIENTS FOR BODY-ORIENTATION TASKS ($N = 44$)

	Trial 1 with Trial 2	Trial 2 with Trial 3	Trial 1 with Trial 3
Perceptual component	.64	.66	.62
Motor component	.85	.90	.83

motor component is more consistent from trial to trial than the perceptual component. This is in line with previous results (**8**).

Table XIII gives the correlations between stabilometer performance and orientation chair performance, including *r*'s between these and the factors of height and weight for 44 college students.

TABLE XIII

CORRELATIONS BETWEEN STABILOMETER AND ORIENTATION-CHAIR PERFORMANCES

	2 Stab. E.C.	3 B.O. Motor	4 B.O. Per. E.O.	5 B.O. Per. E.C.	6 Ht.	7 Wt.
1. Stabilometer—eyes open	.65	−.01	−.10	−.10	−.65	−.70
2. Stabilometer—eyes closed		.21	−.10	−.10	−.74	−.69
3. Body orientation—motor component			−.18	−.20	−.03	−.12
4. Body orientation—percept. E.O.				−.20	.04	.10
5. Body orientation—percept. E.C.					.05	−.05
6. Height						.85

The body-orientation task, using the lever system to keep the body still in space, is a perceptual-motor function with the motor component greater in importance, whereas in the task of merely indicating the direction of the motion of the body the perceptual component is of greater importance. From the above correlations little or no systematic relation exists between the perceptual and motor components of body orientation or between perceptual performances in the eyes-open and eyes-closed situations.

A fairly high positive *r* exists between the eyes-open and eyes-closed scores on the stabilometer, although on the average there is a great difference between the performances in these two tasks in favor of the eyes-open scores.

Apparently there is no systematic relationship between balancing skill in the standing position and either manual reorientation of the body or perceptual sensitivity of body rotation. The factors of weight and height have no bearing on either manual reorientation or perception of rotation.

Correlations between Stabilometer Performance, Eye-Hand Coördination, Steadiness, and Simple Reaction Time

The adequacy of eye-hand coördination was obtained from *S*'s performance on an improved model of the Pursuit-Oscillator. The old model was described in detail in a previous report (9). The main essentials of the new model are given here. A small bakelite platform (six by eight in.) carrying a silver target (one-half in. in diameter), sunk flush with the surface of the platform, is oscillated right and left through a distance of four in. at a constant rate of approximately one oscillation per sec. by a crank and pulley attached on one axle of a speed reducer. The other axle of the speed reducer is connected to a constant speed motor by a V-belt

A hinged gravity-stylus is connected in circuit with an interrupter-commutator, an electric counter and the platform target. *S*'s task is to stand before the oscillating platform and to hold the stylus on the oscillating target. If *S* is successful in holding the stylus on the target the counter will score continuously. The trials are one min. in length with rest periods of one min. A perfect score is 510 per min. *S*'s height is adjusted by using several platforms of various heights to permit a standard distance of three in. between the lower level of *S*'s elbow and the platform level. This type of performance is definitely a learning function.

Steadiness was measured by a new apparatus which controls the rate of movement of *S*'s hand and arm. In view of Gurnee's experiment (10) with the S-tracer with which he found the rate of movement difficult to control, the author decided to retain, in the main, the excellent features of the S-tracer and in addition control the rate of movement. The rate of movement was controlled in the following manner. An S-groove three-sixteenths of an in. wide was cut in a steel plate in the horizontal direction. Immediately behind the S-plate another steel plate with a groove three in. long and one-fourth in. wide was mounted on rollers and connected to a crank from a speed reducer in a manner to produce a moving aperture. The groove in the second plate was perpendicular to the general direction of the S-groove. The shape of the moving aperture naturally varies as the groove in the second plate moves along the S-groove. The rate of movement throughout the stroke varies in accordance with the harmonic-motion produced by the crank on the speed reducer. A constant speed motor is attached to the speed reducer by a V-belt. The time for one complete stroke or oscillation of the aperture is 13 sec. or approixmately 0.9 in. per sec. The rate of movement is sufficiently slow to be well within the reaction time of the vast majority of *S*s.

A brass rod one-tenth of an in. in diameter and one and one-half in. long is inserted into a small wooden handle to serve as a stylus. The stylus is in circuit with a relay, the S-plate, and

the vertical groove plate. *S*'s task is to hold the end of the stylus within the moving aperture without touching the edges of the aperture. By the manipulation of a switch either a positive score or a negative score may be obtained. An interrupter on the speed reducer is in circuit with an electric counter. When the switch is in the 'positive score' position and the motor is turned on, the counter scores continuously if the stylus does not touch the edge of the moving aperture. If the stylus touches the edges of the aperture a relay is actuated and the counter circuit is interrupted. With the switch in the 'negative score' position the counter is actuated only when the stylus touches an edge of the aperture.

Simple reaction time was measured by a simulated motor-car setup. *S* sat with his hands on the steering wheel directly in front of which two neon lights were mounted. *S*'s task was to move the wheel as quickly as possible to the right when the stimulus light on the left was illuminated and to the left when the stimulus light on the right was illuminated. When *E* pushed the control button thus illuminating one of the stimulus lights, a Cenco impulse counter was actuated and continued to operate until *S* turned the steering wheel in the appropriate direction interrupting the counter circuit. The reaction time was recorded in 1/120 sec. Ten consecutive trials for the stimulus light on the right were followed by 10 consecutive trials for the light on the left. The time interval between stimuli was varied to prevent the development of a rhythmic response pattern. The average for the 20 trials was taken as the reaction time for each *S*.

Table XIV gives the correlations between postural balance skill, manual pursuit, manual steadiness and simple reaction time.

TABLE XIV

CORRELATIONS BETWEEN SCORES ON STABILOMETER, MANUAL STEADINESS, MANUAL PURSUIT, AND REACTION TIME

	2	3	4	Reliability Coefficients
1. Stabilometer	.56	−.16	.14	.90 (consecutive trials)
2. Manual steadiness		.52	−.32	.77 (trials 2 & 3—trials 4 & 5)
3. Manual pursuit			−.30	.93 (trial 7—trial 8)
4. Reaction time				.60 (odds—evens) (20 consecutive trials for 47 Ss)

It is clear from Table XIV that there is practically no correlation between stabilometer performance and manual pursuit or between stabilometer performance and simple reaction time. However, there seems to be a steadiness factor in the two skills of balancing and eye-hand coördination as measured by our techniques. Furthermore, the speed factor seems relatively unimportant in manual pursuit and steadiness as well as in balancing skill.

SUMMARY

This paper reports an experimental analysis of various components of dynamic and static equilibrium and a number of correlations among these components.

Dynamic equilibrium was studied in two quite different experimental situations. First, the adequacy of balancing skill with and without visual cues and after mild exercise was measured by an unstable platform technique which presented a real challenge to a standing person. Secondly, perceptual and motor components of body

orientation were studied by a new double-axle rotation chair which presented accelerated movements of different time-lengths and magnitudes. Body sway in the standing position (a static equilibrium component) was measured by the Miles ataxiameter at the head level in cm., with and without visual cues, with normal respiration, and during short periods of total repression of respiration, and after mild exercise. Measurements of weight, height, foot length, and CG-distance were recorded and correlated with dynamic and static equilibrium components.

To determine the relationships between static and dynamic equilibrium and other motor skills involving primarily voluntary responses, steadiness of manual movement, eye-hand coördination and simple reaction times were recorded.

The major conclusions may be summarized as follows:

1. The dynamic component of equilibration, characterized as an orientating perceptual-motor adjustment of the body while in motion in relation to gravity, is quite *unrelated* to the static component, characterized as a continuous tonic reaction in maintaining fixed positions of the body in relation to gravity. This unrelatedness is indicated in the practically zero correlation between performance on the stabilometer and the amount of body sway at the head in the standing position.

2. There is practically no correlation between balancing skill on the stabilometer and ability to maintain manually orientation of the rotation chair which is continually being displaced. Presumably, these two functions are different aspects of dynamic equilibrium.

3. The low correlations between perceptual and motor components of body orientation on the rotation chair suggest a striking disparity between perceptual sensitivity to passive body motion and voluntary motor adjustment.

4. Relatively, weight is of great importance and height of little importance in dynamic stabilometer performance as determined by controlling height and varying weight, likewise by controlling weight and varying height. Weight and height have no bearing on sway scores, rotation scores, steadiness and manual pursuit. The distance from the center of gravity to the foot base line has high correlation with standing height. CG-distance/wgt. has high correlation, whereas CG-distance/hgt. has low correlation, with stabilometer score. Thus, again, weight is the important factor in balance of this type.

5. A small sex difference was found in favor of the women in stabilometer performance with weight controlled.

6. Both static and dynamic equilibrium are aided greatly when visual cues are present; the finer the visual points of reference the better the performance.

7. Mild exercise has little effect on dynamic stabilometer performance, but increases body sway significantly due to increase in respiration which in turn increases the head movements.

8. There is evidence for assuming the presence of a steadiness factor in balancing skill and in eye-manual coördination as indicated by correlations in the order of .50.

(Manuscript received November 10, 1944)

References

1. Travis, R. C. A new stabilometer for measuring dynamic equilibrium in the standing position. *J. exp. Psychol.*, 1944, **34**, 418–424.
2. Hellebrandt, F. A., Riddle, K., & Fries, E. C. The influence of postural sway on stance photography. *Physiother. Rev.*, 1942, **22**, 88–96.
3. Hellebrandt, F. A., Riddle, K., Larsen, E. M., & Fries, E. C. Gravitational influences on postural alignment. *Physiother. Rev.*, 1942, **22**, 143–149.
4. Miles, W. R. Static equilibrium as a useful test of motor control. *J. industr. Hyg.*, 1922, **3**, 316–331.
5. Fearing, F. S. The factors influencing static equilibrium. An experimental study of the effects of practice upon amount and direction of sway. *J. comp. Psychol.*, 1924, **4**, 162–183.
6. Edwards, A. S. The measurement of static ataxia. *Amer. J. Psychol.*, 1942, **55**, 171–188.
7. Harris, J. A., Jackson, C. M., Paterson, D. G., & Scammon, R. E. *The measurement of man.* 1930, p. 49.
8. Travis, R. C. Perception and bodily adjustment under changing accelerations of rotary movement: a new technique. *Amer. J. Psychol.*, 1944, **57**, 468–481.
9. Travis, R. C. A manual pursuit-oscillator. *Amer. J. Psychol.*, 1938, **51**, 406–409.
10. Gurnee, H. Measuring the steadiness of manual movement. *Amer. J. Psychol.*, 1939, **52**, 274–276.

APPARENT WHITENESS IN RELATION TO ALBEDO AND ILLUMINATION *

BY ROBERT E. TAUBMAN

Columbia University

In a recent publication on whiteness constancy, Hsia (1, pp. 48 ff.) emphasizes the importance of formulating an expression of the relation between apparent whiteness, on the one hand, and albedo and illumination, on the other—an expression of *judged whiteness* as it is related to the physical measures of albedo and illumination. The research reported here is an attempt to study this relation.

Apparatus and Procedure

A modification of the apparatus designed by Hsia (1, pp. 20–26) was used in this investigation. It consists of a beaver-board box (95 in. high, 45 in. wide in front, 40 in. high in back, and 24 in. deep) set on a large table. The box is partitioned into left and right chambers open in front and rear. These chambers serve as observation fields. With the upper part of these stimulus boxes covered by black cloth, observation fields 15 in. high in front and 14 in. in back were available.

In this study, only the box to the left of *S* was used. Its inside walls and floor are covered with dull gray paper sheets (albedo equal to .17). This surface prevents the color of the beaver-board or floor of the observation field from interfering with the quality of the illumination or the stimulus proper.

An iron rod stimulus holder, one cm. in diameter and painted dull black, was attached to the rear of the stimulus chamber. The rod was arranged to extend first toward the background, then downward and finally in toward the chamber. On the tip of the rod, an aluminum disc-platform three in. in diameter was attached at an angle of 80 degrees. Two small brass teeth were affixed to the bottom of this platform to hold the stimuli in place. The center of this platform was approximately three in. away from the rear of the stimulus box, eight in. from the center partition and 13 in. from the side wall.

The stimuli used were Hering gray papers. They were mounted on aluminum discs slightly larger than the platform on which they were placed for observation. Gloves were worn by *E* to protect the stimuli from becoming soiled in handling.

The illumination was provided by a frosted 100-watt bulb within the stimulus box; as described by Hsia, this bulb is attached to sashcords and pulleys on the outside of the box to enable *E* to change the illumination by raising or lowering the lamp to the desired levels. These adjustments were made according to a scale in half-centimeters arranged on the outside of the chamber.

S sat in a low chair directly in front of the stimulus chamber, with the stimulus disc at about eye level. *E* was hidden from view by black drapes which extended from the side of the stimulus chamber about 5½ feet towards the rear and then to a near wall. These drapes combine with the darkness in the room to form a background against which the stimuli appear. *E* placed the proper stimulus plate on the holder, set the desired level of illumination, and told *S* to proceed at his convenience. When ready, *S* pressed a manual switch with which he was provided. In the 'on' position, the switch put on the lamp in the stimulus compartment. In the 'off' position, the switch put on a desk lamp (shielded from view by the above-mentioned drapes) which allowed *E* sufficient light to adjust both the stimuli and illumination from trial to trial. In this latter position, the switch also turned on a light in the chamber to *S*'s right. Thus, illumination was provided *S* in which to record his judgments and which at the same time prevented prolonged dark adaptation.

* This problem was undertaken for an advanced class in Experimental Psychology conducted by Prof. R. S. Woodworth, assisted by Dr. E. A. Jerome. The author is indebted to these men for advice and suggestions concerning this study.

S was given the following instructions:

> You are familiar with the statement that a white object in shadow may appear white, in spite of the fact that it is not reflecting much light. Hence, it has been said that whiteness is not the same as brightness. It is possible, however, that the appearance of the object would be different under other illuminations. We are making a study of whiteness under different levels of illumination.
> You will be shown a series of grays, containing from little to much whiteness, each stimulus appearing under different levels of illumination. Think of a pure white and a pure black as end anchors of the stimulus scale. Place the stimuli in this study on a scale with such end anchors and then judge each stimulus in terms of its percent whiteness on a scale from zero to 100 percent.

S was allowed to adapt for about five min. to the general illumination in which he was to observe. After this period, he ran through five practice trials with the entire series of stimuli and lights, both of which were varied randomly. At the end of this series, *S* was given a rest and the data reported here were taken. As much time as was needed was given for making the judgments. Rests were allowed only after a whole series of stimuli had been finished. From four to six hours were required for the experiment, extending usually to two successive days. Where a second day was required, *S* had a practice trial which again consisted of an entire series of stimuli under the different light levels. This was done to allow *S* to pick up his accustomed scale.

In Part I, there were six grays and the albedo intervals among them are quite equal, except for the highest step for which a more satisfactory albedo could not be obtained.[1] Part II was an attempt to improve the spacing among both the levels of illumination and the stimuli. The grays are in approximate ratio of 1 : 2, whereas the levels of illumination are in the ratio of 1 : 1.66. In this second part, the number of grays was reduced to four, though the stimulus range was kept intact, whereas the range of illumination was increased.

Subjects

Four well-trained *S*s (graduate students in psychology) were used. One of these, SS, observed in both parts of the study.

Results

All the data are presented in the following tables.

TABLE I

Data for Part I

The figures within the table are Mean Judged Whiteness (expressed in percent) based on 10 judgments of each stimulus at every level of illumination.

	Level of Illumin.*	Albedo					
		.07	.10	.14	.19	.26	.54
S-SS	1.20	10	16	32	53	63	76
	2.00	13	28	44	57	69.5	76
	3.77	16	36	49	61	72	79.5
	8.50	10	19	36	58	66	78
S-EAJ	1.20	20.3	27	32	70	81.5	87.6
	2.00	23.5	40.5	58.8	73.3	84.1	94.6
	3.77	38.5	57	69.4	78.3	88.4	96.6
	8.50	34	47.5	62.5	77.7	85.9	97.1

* Level of illumination is in ft.c. taken from a theoretical reference plate of 100 percent reflection.

[1] All measurements were made with the Macbeth Illuminometer.

TABLE II

DATA FOR PART II

The figures within the table are Mean Judged Whiteness (expressed in percent) based on 10 judgments of each stimulus at every level of illumination.

	Level of Illumin.*	Albedo			
		.07	.14	.26	.54
S-SS	1.20	11	24	60	78
	2.00	14	24	65	79
	3.32	24	55	68	80
	5.51	35	48	70	80
	9.15	11.5	35	61	79
S-TK	1.20	22.5	25.5	44	58
	2.00	27	35.5	47.5	61.5
	3.32	34.5	38	52	66.5
	5.51	38	39.5	59	69
	9.15	33.5	40	55	69.5
S-BKS	1.20	16	21.5	55	85
	2.00	18.5	23.5	62	87
	3.32	25	27	66	89.3
	5.51	26	34	71	90
	9.15	26	40.5	68.5	90

* Level of illumination is in ft.c. taken from a theoretical reference plate of 100 percent reflection.

DISCUSSION

The data in our experiments may be analyzed in two ways. First, we may study the single level of illumination and observe the rise in the judged whiteness function as we go up the albedo scale; here, we hold the illumination constant and vary the albedo presented. Second, we may examine the single albedo and observe the judged whiteness function as we increase the illumination present; this is to say, we hold the albedo constant and go up the illumination levels.

Our data in Part I, for subjects EAJ and SS, indicate that when we hold the illumination constant and go up the albedo scale, there is a rise in the judged whiteness function. Further, it may be seen that the judgment curve is negatively accelerated. Some heterogeneity is evident in the data, but the essential form shows negative acceleration.

If we now hold a given stimulus gray constant and observe the obtained function as we increase the illumination, we find highly significant material revealed therein. First, it is clear that judged whiteness increases for the lower albedo values (.07, .10, .14) as a function of greater illumination; and second, there is a drop in judged whiteness at the highest level of illumination. The objective situation here is such that the albedo value remains constant whatever the amount of illumination. Yet, up to a point, there is an increase in apparent

whiteness with each illumination increment; and beyond this point, an increase in illumination is marked by a decrease in apparent whiteness. The reasons for this obtained fall are not clear. In any event, the data thus indicate that apparent whiteness is directly related to physical brightness, but that this is true only to a certain degree—at an extreme illumination, the relation breaks. With respect to the higher albedo values (.19, .26, .54) we note that here there is only a slight tendency for an increase in judged whiteness and then a fall at the highest illumination step. The data demonstrate that the three lightest grays show the greatest amount of constancy. Thus, we may say that, for these higher albedos, a change in illumination does not bring about as much change in apparent whiteness as for the lower albedos.

Our analysis here also demonstrates the presence of a well-differentiated series of stimulus grays, with no overlapping of judgments present. Finally, we may point out that there are individual differences in the observers' use of the judgmental scale.

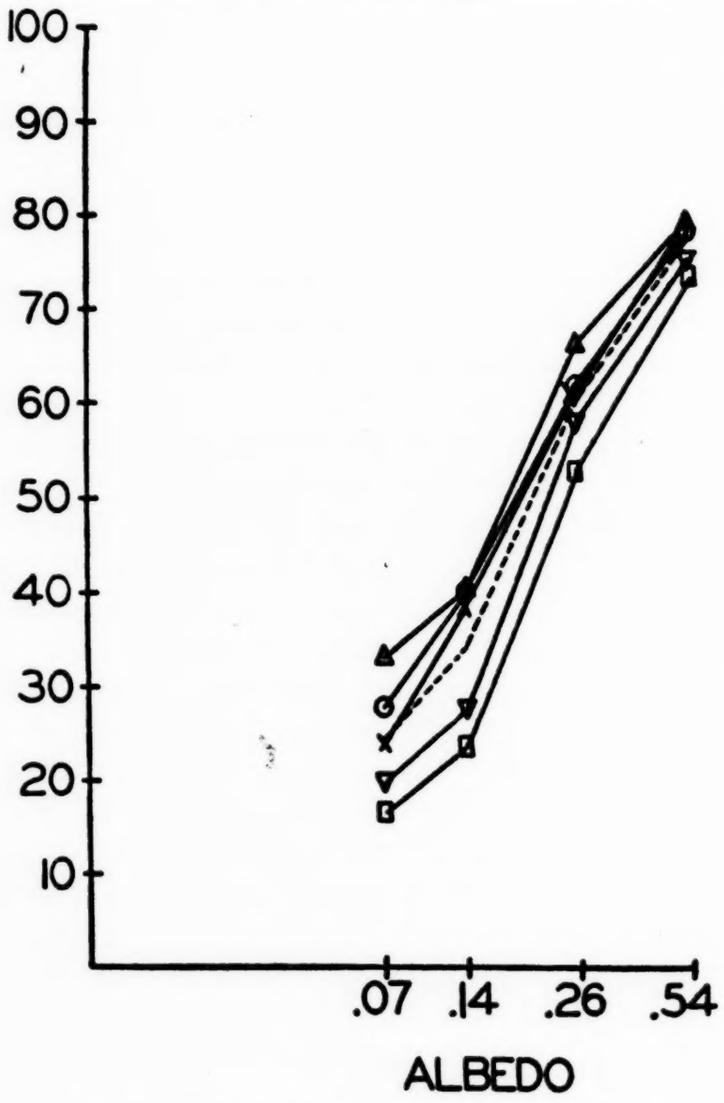

FIG. 1. Judged whiteness as a function of albedo

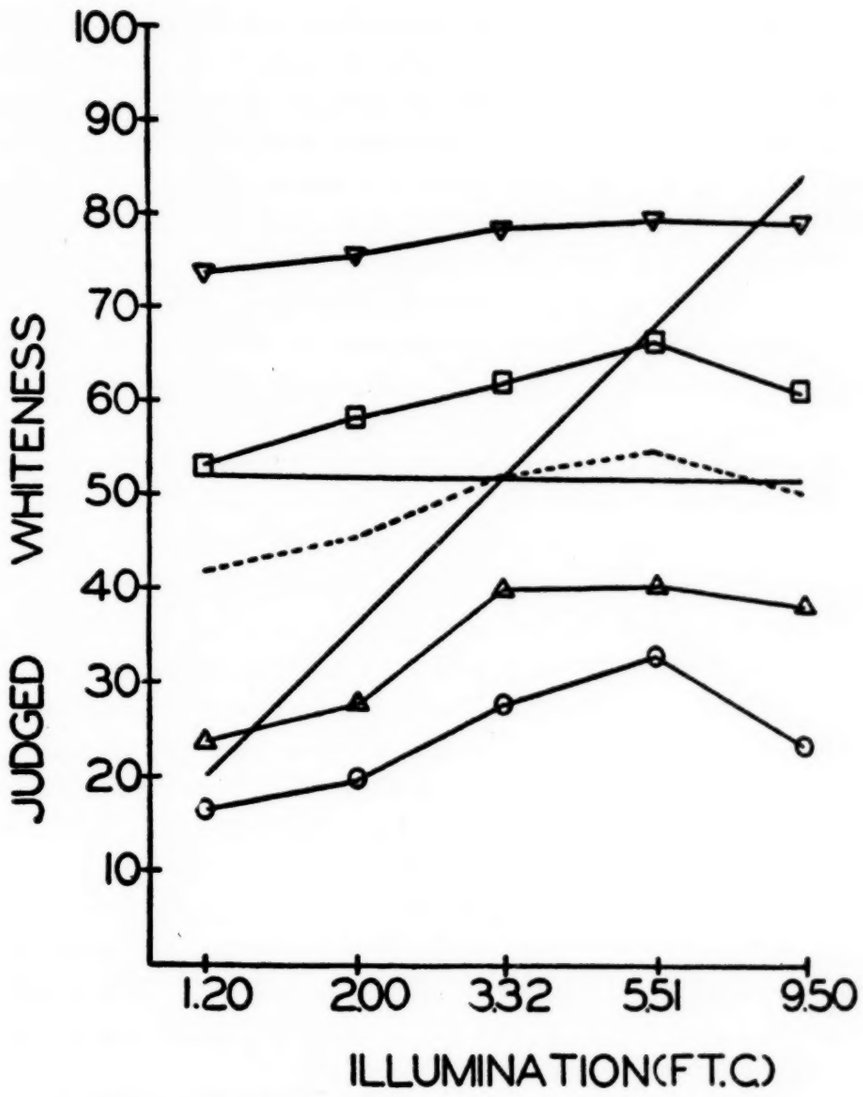

FIG. 2. Judged whiteness as a function of illumination

Figs. 1 and 2 are derived by averaging the data presented in Table II. Fig. 1 represents judged whiteness (plotted arithmetically) as a function of albedo (on a logarithmic scale), whereas Fig. 2 shows judged whiteness (plotted arithmetically) as a function of illumination (on a logarithmic scale). (See text for full description of these plots.) The following symbols have been used.

FIG. 1

Symbol	Illumination Level
□	1.20
▽	2.00
○	3.32
△	5.51
×	9.15

FIG. 2

Symbol	Albedo
○	.07
△	.14
□	.26
▽	.54

It is to be remembered that our improved spacing of the stimulus grays and levels of illumination in Part II allows us to test the kind of curve obtained more adequately. Figs. 1 and 2 are the graphic representations of the data in Part II made on semi-logarithmic coördinates. They are average curves derived from the data of the three Ss studied.

When we plot judged whiteness arithmetically against albedo on a logarithmic scale (Fig. 1) we observe a family of curves resembling ogives. There is considerable rise in judged whiteness over the albedo range, but the data evidence discontinuity between albedo values of .14 and .26. It seems that the two lower albedos were judged as near-blacks and the two higher albedos as near-whites. Thus, it would seem, according to our findings, that our *Ss* were not judging along a continuous black-gray-white scale, but in two large categories—black and white, without any reference to the middle of the scale. Actually, the introspections of our *Ss* indicated great difficulty in locating or using the middle scale values. Common life experience with albedos such as we have used supports the black-white dichotomy observed here.

Fig. 2 shows judged whiteness plotted arithmetically against illumination on a logarithmic scale. In general, it is clear that these curves show less change than those of Fig. 1; thus, considerable constancy is shown here. To a degree, however, judged whiteness does increase as a function of illumination. The two lower albedos show a greater rise, or less constancy, while the two higher albedos are less affected by illumination changes and thus manifest greater constancy. Finally, it should be pointed out that the interesting rise and fall mentioned above are in evidence here also.

Fig. 1 presents average judgments for all illuminations, indicated by the broken line curve. Fig. 2 likewise presents average judgments for all albedos, indicated by the broken line curve. It is evident that these mean curves are faithful representations of their respective families of curves.

The question may be considered: How much constancy is shown in the data? When albedo is changed, judged whiteness should keep pace with it. Hence, the albedo graph may be used to show how much change in judged whiteness should be produced by changed illumination (1) if constancy were perfect, that is, if apparent whiteness was completely determined by albedo and independent of illumination; and (2) if constancy were zero, that is, if apparent whiteness was directly related to stimulus brightness. In the latter case, doubling illumination should have the same effect as doubling albedo. Now, taking the points on Fig. 1 where the broken line crosses an albedo value equal to .1 and an albedo equal to .4 (so taking a fair representation of the lower albedos and of the higher ones), it will be found that the mean judged whiteness is 30 for albedo .1 and 70 for albedo .4; this represents a difference of 40 points for doubling and redoubling, or 20 points for doubling once. Thus, it can be seen that doubling albedo raises judged whiteness by 20 points, considering the middle of the scale. Applying this result to Fig. 2, we may take the

middle point of the mean judged whiteness, that is, the point at illumination equal to 3.32, and draw an oblique line with such a slope as will produce a change of 20 points judged whiteness for doubling or halving the illumination. If changes of illumination had the same effect as changes of albedo, as they have in respect to the stimulus actually received from the disc, the illumination mean curve would be this oblique line. Zero constancy would be represented by this oblique line, whereas the horizontal line would represent 100 percent constancy.

Summary

Two experiments were done to study the relation obtaining between apparent whiteness and its physical correlates, albedo and illumination. Hering grays served as stimuli in an apparatus designed by Hsia (1). There was found a close relation between apparent or judged whiteness and illumination at low albedo values, whereas the apparent whiteness of the higher albedos is only slightly affected by illumination changes, thus resulting in high constancy. Our data evidence a break in the apparent whiteness-albedo function. This means that albedo was judged in the two large categories of black and white, rather than along a black-white continuum. Finally, the form of the judged whiteness function is indicated under conditions of perfect, as well as zero, constancy.

(Manuscript received November 24, 1944)

References

1. Hsia, Y. Whiteness constancy as a function of difference in illumination. *Arch. Psychol., N. Y.*, 1944, No. 284.

SELECTIVE FORGETTING WHEN THE SUBJECT IS NOT 'EGO-INVOLVED'

BY FRANKLIN J. SHAW

University of New Hampshire

AND

ALICE SPOONER *

Indiana University

In a previous article (3) Shaw presented evidence supporting the hypothesis that (1) agreement and disagreement with an opinion held by the *S* and (2) favorable and unfavorable evaluations of the *S* are determinants of selective forgetting. A favorable evaluation was defined as the attributing of a desirable trait to an *S* by an alleged rater or the failure to attribute an undesirable trait to the *S* whereas an unfavorable evaluation was defined as attributing an undesirable trait or not attributing a desirable trait. In the present investigation no such evaluations were made of the *S*s, but they were made of individuals who had been previously rated by the *S*s. Thus, under the conditions of this experiment, the *S*s were not 'ego-involved.' It is our hypothesis that under these conditions, favorable and unfavorable evaluations will not have a differential effect upon recall, but that agreement and disagreement between the *S*'s ratings and alleged ratings will result in selective forgetting.

PROCEDURE

The procedure is a modification of one originally used by Wallen (4) and later by Shaw (3). The details follow.

Seventeen *S*s, all of whom were undergraduates, were given a list of 45 adjectives and asked to rate some specified person with whom the *S* was well acquainted. The name of the person appeared at the top of the page on which the adjectives were printed with the instruction to rate that person. The rating was done by marking a plus beside the adjective if the *S* thought it applied to the specified individual and a zero if it did not.

One week later the *S*s were given the same list of adjectives with pluses and zeros beside them. They were told that the pluses and zeros represented composite ratings of the individual whom they had rated a week previously. It was said that the composite ratings represented the judgments of several individuals who knew the rated person. Actually these ratings were not genuine but were made up by changing 23 of the 45 ratings which the *S*s had made of the specified person in accordance with a pre-arranged scheme. The *S*s examined the bogus composite ratings, which were presented as genuine, while *E* read over the list of adjectives twice at a rate of three or four sec. per word.

* This study was conducted by Miss Spooner, who is now a graduate student at Indiana University, at Bates College under the direction of the senior author.

A week after the *S*s had examined the alleged composite ratings in the manner described in previous paragraph, they were asked to recall these ratings. As soon as the recall was completed, the *S*s were instructed to mark each adjective as representative of a desirable or undesirable trait by marking a *D* or a *U* after the adjective.

Results

The symbols employed in reporting the results have been used previously by Wallen (4) and Shaw (3). A *D* indicates that the *S* considered the trait desirable and a *U*, undesirable. The sequence of the symbols, x and o, shows how they were marked by the *S* and in the alleged rating respectively. Thus, the combination of symbols, Dxx, stands for traits which the *S* considered desirable and which both he and the alleged 'rater' considered descriptive of the rated person. The combination, Dxo, stands for traits which the *S* considered desirable and which he considered descriptive of the rated person, but which the alleged 'rater' did not consider descriptive of the rated person. Table I contains the items of each type, their total number, and the percentage of errors made in recall of each type.

TABLE I

The Total Number of Each Type of Item and the Percentage of Errors Made in Recall of Each Type

Type	Total	Percentage Wrong
Dxx	154	26.0
Doo	79	31.6
Uxx	62	21.0
Uoo	79	31.6
Dxo	148	56.8
Dox	81	49.4
Uxo	46	39.1
Uox	105	51.4

In Table II data are presented showing the influence of favorable and unfavorable evaluations, contained in the alleged composite rating, of the rated person upon recall. A favorable evaluation consisted of an assignment in the alleged rating of a trait considered desirable by the *S* or the failure to assign a trait considered undesirable by the *S*. An unfavorable evaluation consisted of assigning an undesirable trait or not assigning a desirable trait. In Table II, the influence of agreement or disagreement between the *S*'s ratings and the alleged composite ratings is held constant.

It will be recalled that the hypothesis set forth in the introduction was that under the conditions of this experiment in which the *S*'s personality was not 'evaluated,' favorable and unfavorable evaluations would not have a differential effect upon recall. The small chi-square values in Table II lend support to this hypothesis. The differences between percentages of errors in recall of favorable and unfavorable

evaluations cannot be considered statistically significant in any of the comparisons.

Although, according to our hypothesis, favorable and unfavorable evaluations would not have a differential effect upon recall under the conditions of this experiment, it was also postulated that the condi-

TABLE II

THE RELATIONSHIP BETWEEN EVALUATION, CONTAINED IN THE ALLEGED RATING, OF THE INDIVIDUAL PREVIOUSLY RATED BY THE *S* AND PROPORTION OF ERRORS IN RECALL

Type	Percentage of Errors	Chi-Square
Uox (*u*)	51.4	1.9
Uxo (*f*)	39.1	
Uxx (*u*)	21.0	2.0
Uoo (*f*)	31.6	
Dxo (*u*)	56.8	1.2
Dox (*f*)	49.4	
Doo (*u*)	31.6	.8
Dxx (*f*)	26.0	

The symbol *u* indicates an unfavorable evaluation, contained in the alleged rating, of the rated person.

The symbol *f* indicates a favorable evaluation, contained in the alleged rating, of the rated person.

tions were such as to lead to better recall of those items which agreed with the *S*'s opinion than those which did not. Table III supports this expectation. In this table the influence of evaluation is held constant in the comparisons made. In every case the percentage of errors in recall is smaller when there is agreement. The differences between percentages of errors in recall are significant beyond the one percent level in all except one instance.

TABLE III

THE RELATIONSHIP BETWEEN AGREEMENT AND DISAGREEMENT WITH THE *S*'s OPINION OF THE RATED PERSON AND PROPORTION OF ERRORS IN RECALL

Type	Percentage of Errors	Chi-Square
Uxo (*d*)	39.1	.7
Uoo (*a*)	31.6	
Uox (*d*)	51.4	15.0
Uxx (*a*)	21.0	
Dox (*d*)	49.4	13.0
Dxx (*a*)	26.0	
Dxo (*d*)	56.8	13.0
Doo (*a*)	31.6	

The symbol *d* indicates disagreement between the *S*'s opinion of the rated person and the opinion contained in the alleged rating.

The symbol *a* indicates agreement between the *S*'s opinion of the rated person and the opinion contained in the alleged rating.

In the comparisons made in Table IV the influences of agreement-disagreement with the *S*'s opinion and evaluation are held constant and the possible effect of the *S*'s estimate of desirability or undesirability of a trait is allowed to vary. That desirability and undesirability, as such, have little differential effect upon recall seems to be demonstrated by the data in Table IV. None of the differences between percentages of errors in recall can be considered statistically significant. This finding is, of course, in accord with the accumulation of considerations (**2**) suggesting that pleasantness and unpleasantness are not in themselves determinants of selective forgetting.

TABLE IV

THE RELATIONSHIP BETWEEN THE *S*'s ESTIMATE OF DESIRABILITY OR UNDESIRABILITY OF A TRAIT AND PROPORTION OF ERRORS IN RECALL

Type	Percentage of Errors	Chi-Square
Dxx (*f*) (*a*)	26.0	.8
Uoo (*f*) (*a*)	31.6	
Dxo (*u*) (*d*)	56.8	.7
Uox (*u*) (*d*)	51.4	
Dox (*f*) (*d*)	49.4	1.2
Uxo (*f*) (*d*)	39.1	
Doo (*u*) (*a*)	31.6	2.0
Uxx (*u*) (*a*)	21.0	

The symbol *d* indicates disagreement between the *S*'s opinion of the rated person and the opinion contained in the alleged rating.

The symbol *a* indicates agreement between the *S*'s opinion of the rated person and the opinion contained in the alleged rating.

The symbol *u* indicates an unfavorable evaluation, contained in the alleged rating, of the rated person.

The symbol *f* indicates a favorable evaluation, contained in the alleged rating, of the rated person.

DISCUSSION OF RESULTS

In a previous investigation by Shaw (**3**) in which the *S*s were assigned bogus ratings presented as genuine, it was found that there was a significantly smaller number of errors in recall of those items representing a favorable evaluation than of those representing an unfavorable evaluation. The explanation proposed for this finding was as follows (**3**, p. 443):

> Mowrer's conception of the role of anxiety in learning affords a suggetsion as to why unfavorable evaluations are not recalled as well as favorable evaluations. An unfavorable evaluation is, of course, a form of disapproval which is a potent stimulus to the arousal of anxiety. Mowrer points out that anxiety can be regarded as a motivating condition, the reduction of which constitutes reinforcement or reward. The specific motivation will be getting away from the stimulus evoking the anxiety. Any activity which succeeds in achieving this will be reinforced. This would mean, then, that the perception of an unfavorable evaluation which would accompany anxiety would not be rewarded, but that activity, including perceptual activity, which got away

from the stimulus evoking the perception of an unfavorable evaluation would be rewarded. A favorable evaluation, on the other hand, constitutes secondary reward, and it would, therefore, be expected that favorable evaluations would be better retained.

In this investigation the *Ss* were not rated. This meant, of course, that they were not the objects of favorable and unfavorable evaluations. Under these conditions it would not be expected that an unfavorable evaluation would arouse anxiety which would in turn result in poorer recall of unfavorable evaluations. The findings are consistent with this expectation and lend support, we believe, to the interpretation of the results of the previous study in which the *Ss* who were rated did recall favorable evaluations with fewer errors than they recalled unfavorable evaluations.

In both the previous study and the present investigation, agreement and disagreement with the *S*'s opinion do have a selective effect upon recall. The following was offered as a possible explanation (**3**, p. 442):

> The reason for this may be that we acquire material more easily when it is in line, so to speak, with our opinions and attitudes than when it is not. Thus, a man may assimilate readily those parts of a speech which are in accord with his opinions because they fit in with habitual modes of responding. He may not assimilate as readily those parts which disagree for the opposite reason. It follows that he would recall more readily the portions which agree than which disagree because they were better assimilated in the first place.

If this explanation is correct, the problem remains to account for the better assimilation of material in line with an opinion.

A possibility that may merit consideration is that an existing opinion establishes a predisposition to perceive events in accord with the opinion and to ignore events that are not. This seems particularly relevant to the data presented here, since the *Ss* made their own ratings prior to examining the alleged composite ratings and the former could have set up proactive inhibition against the latter. In terms of a competition theory this would mean that there was conflict when the ratings disagreed but not when they agreed. It would follow that the alleged ratings that agreed with the *S*'s ratings would be recalled better.

That an existing opinion may predispose toward perceiving events in accord with the opinion is shown by an investigation by Zillig (**4**). The concept of proactive inhibition also seems to apply to her results. After arranging to have disliked children perform exercises more skillfully than liked children, she asked other children to judge the performances of the two groups. According to the judgments, the performance of the liked children was superior. Zillig reports that the judgments apparently were not deliberately dishonest, but that the children who judged actually perceived the performance of the liked group as superior. It seems plausible that the habitual manner of

judging the children in the disliked group was in conflict with more favorable judgments or that what might be called proactive inhibition against more favorable judgments was sufficiently strong to prevent their appearance.

Summary

In a previous study conducted by Shaw (2), it was found that agreement and disagreement with an opinion held by the *S* and favorable and unfavorable evaluations of the *S* had a differential effect upon recall. In that study, the *S*s were asked to recall bogus ratings of themselves presented as genuine. In this investigation, on the other hand, *S*s were asked to recall bogus ratings, presented as genuine, of individuals whom they had previously rated. It was our hypothesis that under these circumstances agreement and disagreement with the *S*'s opinion would have a selective effect upon recall but that evaluation would not since the *S*'s own personality was not involved in the ratings the subject was required to recall. The results supported this hypothesis.

Better recall of items representing agreement with the *S*'s opinion was explained in terms of better acquisition of material which is in accord with an existing opinion. A possible explanation of such acquisition was proposed. The fact that favorable and unfavorable evaluations did not have a selective effect upon recall was accounted for in terms of unfavorable evaluations not arousing anxiety when these evaluations were made of others, rather than of the *S*s.

(Manuscript received November 13, 1944)

References

1. Edwards, A. L. The retention of affective experiences—a criticism and restatement of the problem. *Psychol. Rev.*, 1942, **49**, 43–53.
2. Shaw, F. J. Two determinants of selective forgetting. *J. abnorm. soc. Psychol.*, 1944, **39**, 434–445.
3. Wallen, R. Ego-involvement as a determinant of selective forgetting. *J. abnorm. soc. Psychol.*, 1942, **37**, 20–40.
4. Zillig, M. "Einstellung und Aussage," *Z. Psychol.*, 1928, **106**, 58–106.

AUTONOMIC RESPONSE IN POSTHYPNOTIC AMNESIA

BY M. E. BITTERMAN AND F. L. MARCUSE

Cornell University

Introduction

Degree of posthypnotic amnesia is usually estimated from the results of recall, recognition and relearning tests. Since it is generally recognized that hypnotic procedure can less markedly influence involuntary levels of response (**4**, **12**), it may be asked whether autonomic indices would provide a more sensitive criterion. It was to test this possibility that the present experiment was performed.

Procedure

The *S* was an undergraduate woman student in whom deep hypnotic states had been induced on many previous occasions. She will be referred to as *S*, and the experimenters as E_1 and E_2.

E_1 drew up a list of eight common nouns and selected one of the middle six by lot.[1] He then hypnotized *S*[2], who knew nothing of the nature of the experiment, and gave the following instructions: "I am going to mention a word to you. The word is (here the critical word is inserted). When you are awakened, you will not remember that I mentioned this word to you. It will not be more familiar to you than any other word. No matter how hard you try, you will be absolutely unable to remember that I mentioned this word to you." (Repeat.)

E_2, who did not know the suggested word, questioned *S* after she was 'awakened.' He asked whether she could remember anything that had occurred and, specifically, whether she could recall any word that might have been mentioned to her. Then he read, one by one, the list of eight words and asked whether *S* recognized any of them as having been mentioned. *S*, after some thought, invariably replied in the negative.

After a 48-hour interval, during which time the word was not recalled, *S* returned to the laboratory. E_2 read the list of words once more and at the same time took physiological recordings (respiratory and cardiovascular) on a Keeler Polygraph. Sample records are presented in Fig. 1. A mark was made on the baseline with a signal-magnet as each word was read. *S*'s replies were in the negative. *S* was then rehypnotized by E_1, and, in response to his question, was able to recall the suggested word.

Several days later *S* returned to the laboratory and the procedure was repeated with a second list of words, the roles of E_1 and E_2 being reversed. Work was continued until eight tests had been made, each *E* playing each role four times. The procedure was always the same, except that each *E* selected (by chance) either his second, third or fourth test as a control. In these instances *S* was hypnotized, but no word was suggested to her.

At the conclusion of the tests, each *E* examined privately the four polygraph records which he had taken. On the basis of the criteria ordinarily employed in detecting deception (**6**), he attempted to determine which of the four records was a control and which word on each of the others was the one which had been mentioned in hypnosis.[3] If it happened that *E* was not

[1] This procedure ensured the occurrence of neutral items in the initial and final position of the series and was employed to facilitate the interpretation of the physiological recordings (*vide infra*).

[2] Depth was estimated as 20 on the Friedlander-Sarbin scale (**3**).

[3] The writers had used the 'lie detector' with considerable success on many previous occasions.

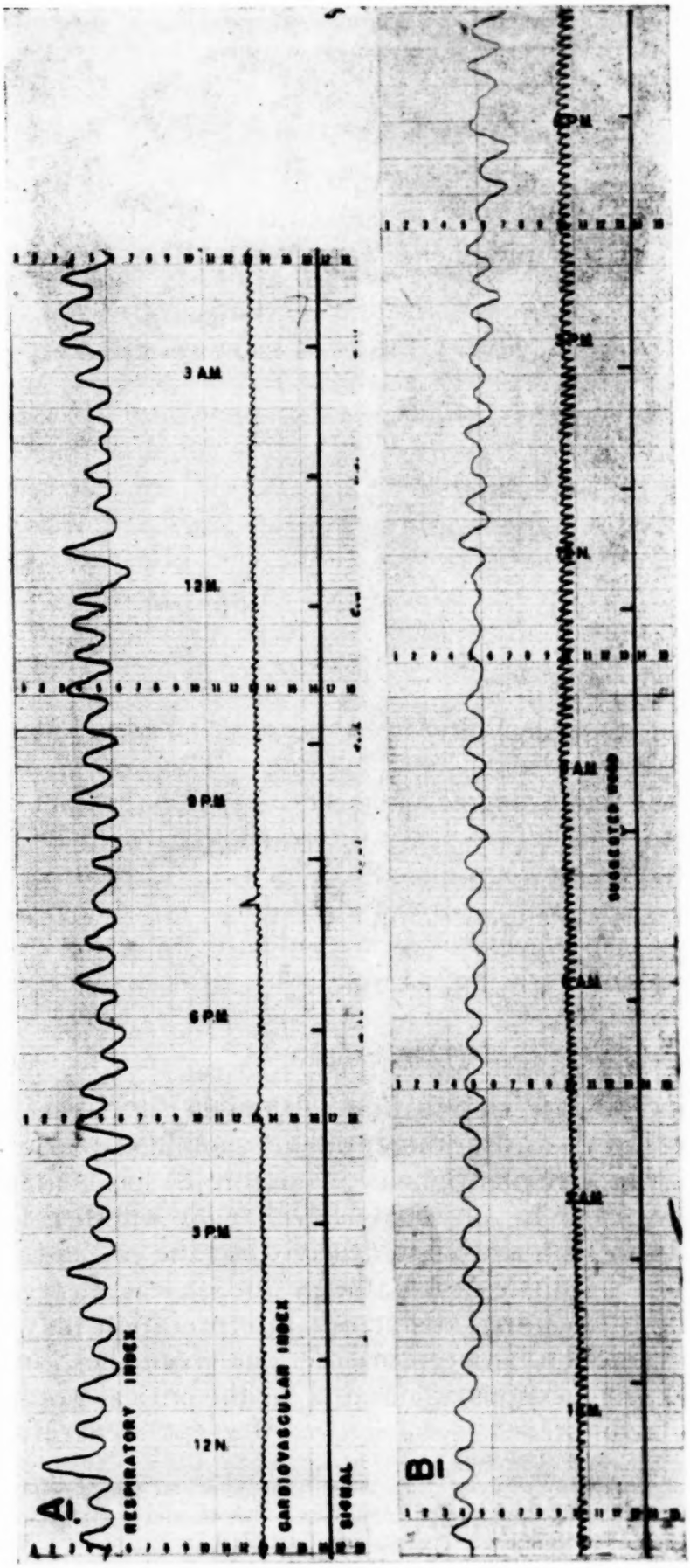

FIG. 1. Records of physiological responses during tests of recognition. *A*. Test 4 (control); *B*. Test 5.

completely certain of his interpretation, he recorded a 'second choice' along with his primary judgment. After the eight decisions had been made in writing, they were compared with notes of what actually had occurred.

Results

It may be seen from Table I that five of the eight primary judgments were correct. Four of these records (1, 4, 5, 6) were sufficiently clear that alternative judgments were not made. In the remaining

TABLE I

Validity of Judgments Based on Polygraph Records

E	Record No.	Suggested Word	E's Primary Judgment	E's Second Choice
1	1	clip	clip	—
	3	dictionary	control day	dictionary
	5	bottle	bottle	—
	7	(control day)	ribbon	control day
2	2	goal	goal	dust
	4	(control day)	control day	—
	6	pole	pole	—
	8	cover	content	cover

three cases the second choices were correct. The results were significantly better than chance ($P < .01$).[4] The two *E*s scored approximately the same degree of success despite the fact that they disagreed as to whether or not positive results would be obtained.[5]

Sample records are reproduced in Fig. 1. 'A,' record 4, is a control, and is characteristically free of differential physiological changes. In 'B,' record 5, the suggested word is indicated by a 'peak of tension' in the cardiovascular tracing, as well as by local changes (**6**).

Discussion

The results indicate clearly that, although the *S* was unable to recognize the critical words, differential autonomic responses occurred when these words were mentioned. Marston (**8**) has contended that deception is revealed by physiological changes whether or not the individual is aware of a motive to deceive, but the concept of 'unconscious deception' has little explanatory value, at least in regard to the present findings. A more satisfactory interpretation may be based upon the contradiction between instructions given the *S* in the hypnotic states, i.e., suggestions of amnesia for the critical words, and the post-hypnotic requests to make every effort at recognition. It is

[4] The writers are indebted to Professor T. A. Ryan for statistical advice.

[5] This result is significant from the point of view of those who stress the importance of operator attitude (**2**).

reasonable to assume that this procedure gave rise to a motivational conflict which tended to produce emotion.[6] This hypothesis appears to be borne out by the absence of marked physiological changes in the control records, and from the same point of view it might be anticipated that, if the experimental procedure were modified to the extent that the *S* is not pressed for recall or recognition of the critical words, no differential autonomic response to the reading of the lists would occur.

There has been considerable controversy in the literature concerning the degree of dissociation which can be obtained by hypnotic procedures. Hull (**4**) and White and Shevack (**12**) have minimized the possibility of producing complete dissociation, while Wells (**9**, **10**, **11**) has argued that with appropriate instruction and sufficient 'depth' of hypnosis, the absolute functional independence of dissociated processes may be achieved. Erickson (**1**) has found it necessary to induce 'profound' trances in order to ensure the freedom of posthypnotic behavior from the influence of waking experience.

In the present experiment complete insulation was not achieved, as evidenced from the interaction of the instructions given in the hypnotic and waking states. If in further research it is found that the physiological indicators reported here do not appear with the use of a 'profound' trance or with suitable modifications of the hypnotic instruction, the possibility of complete dissociation is strengthened, and the technique of the present experiment may be employed as a sensitive test of the degree of insulation which is achieved in hypnotic experiments. If, on the other hand, these indicators continue to appear, the results might be taken as evidence against the possibility of obtaining complete dissociation.

Summary

Amnesia for words mentioned in the hypnotic state was suggested to a somnambule. Despite the *S*'s inability to recognize or recall these words in the posthypnotic period, they could be determined from records of physiological changes which were made during tests of recognition. The relation of this finding to the concept of dissociation is discussed.

(Manuscript received November 17, 1944)

[6] In this connection *S*'s behavior during the course of investigation is significant. She complained of dizziness following the experimental sessions and of insomnia, and she frequently failed to keep appointments. These reactions were not characteristic of *S*, but resembled behavior observed by Luria (**7**) and by Huston, Shakow and Erickson (**5**) in studies of hypnotically induced conflict. In the final hypnotic session and in the subsequent waking period the nature of the experiment was fully explained to *S*, and one week later the symptoms had disappeared.

References

1. Erickson, M. H. Induction of color blindness by technique of hypnotic suggestion. *J. gen. Psychol.*, 1939, **20**, 61–89.
2. Estabrooks, G. H. *Hypnotism.* New York: Dutton & Co., 1943.
3. Friedlander, J. W., & Sarbin, T. R. The depth of hypnosis. *J. abnorm. soc. Psychol.*, 1938, **33**, 453–475.
4. Hull, C. L. *Hypnosis and suggestibility.* New York: Appleton-Century, 1933.
5. Huston, P. E., Shakow, D., & Erickson, M. H. A study of hypnotically induced complexes by means of the Luria technique. *J. gen. Psychol.*, 1934, **11**, 65–97.
6. Inbau, F. E. *Lie detection and criminal interrogaiton.* Baltimore: Williams and Wilkins, 1942.
7. Luria, A. R. *Nature of human conflicts.* (W. H. Gantt, trans.) New York: Liveright, 1932.
8. Marston, W. M. *The lie detector test.* New York: R. R. Smith, 1938.
9. Wells, W. R. Resistance to induced dissociation. *J. abnorm. soc. Psychol.*, 1940, **35**, 261–272.
10. Wells, W. R. Extent and duration of posthypnotic amnesia. *J. Psychol.*, 1940, **9**, 137–151.
11. Wells, W. R. Experiments in hypnotic production of crime. *J. Psychol.*, 1941, **11**, 63–102.
12. White, R. W., & Shevack, B. J. Hypnotism and the concept of dissociation. *J. abnorm. soc. Psychol.*, 1942, **37**, 309–328.